GENETIC ANALYSIS

INTRODUCTION

Limitations of the current approaches

There are a number of limitations to carrying out association studies using single nucleotide polymorphisms (SNPs) and linkage disequilibrium within human populations (see Science, Vol 278, p1580, (1997) for a review of such methods). We have no control over recombination frequency around a given locus or over past human genetic crossing. Some mutations will be closely correlated with nearby SNPs and others will not.

The need for whole genome analysis

With the SNP and linkage disequilibrium approach (and many others), markers are essentially used as a surrogate for sequencing - the more markers, the better. The logical endpoint of the above argument is to look at every base in the human genome - and carry out what could be termed a whole genome association study. In essence, the sequence at every base would be determined for the genome of each member of a phenotypically 'affected' and a phenotypically 'unaffected' population. Statistical correlations (associations) could then be drawn between sequence differences and phenotype. Such associations would have future predictive values for the phenotype of interest, knowing the genotype and could be of great value in medicine and pharmacogenetics.

The current invention selectively enriches for DNA fragments that determine phenotype in the 'affected' population and thus makes the prospect of carrying out whole genome association studies for humans and other species a very real possibility.

Definition of terms used with the current invention

Within the scope of the current invention, the individuals chosen for whole genome analysis may be human, animal or plant and they may be eukaryotic, prokaryotic or archaebacterial in origin.

The terms 'affected' and 'unaffected' are used without limitation in order to categorise individuals into two groups – those that possess a defined phenotype of interest ('affected' individuals) and those that do not possess the phenotype of interest ('unaffected' individuals). The phenotype common to the 'affected' individuals may be either beneficial (e.g. for these individuals, a particular pharmaceutical entity might show high efficacy in a phase II clinical trial) or detrimental (e.g. for these individuals, a particular pharmaceutical entity might show adverse toxicology in a phase I clinical trial).

The 'affected' population may comprise one or more individuals and the 'unaffected' population may similarly comprise one or more individuals according to the particular embodiment of the invention (see below).

The term DNA is used throughout for simplicity. Within the scope of the current invention, the term DNA may equally well apply to all or part of the haploid, diploid or polyploid genomic DNA content of one or more germ line or somatic cell(s). The DNA may be extracted from cells taken directly from the individual(s), the DNA may be extracted from cells cultured or immortalised from the individual(s) or the DNA may be prepared from a library of clones - with inserts derived from the individual(s) and propagated in some appropriate host and cloning vector system. For the particular case wherein the term DNA refers to the expressed part of the haploid, diploid or polyploid genomic DNA content of one or more somatic cells and the DNA is prepared from a library of clones - with inserts derived from the individual(s) and propagated in some appropriate host and cloning vector system, a cDNA library (normalised or otherwise) may be used.

In the current invention, DNA is compared in fragmented form. Fragmentation can be performed after DNA extraction, prior to cloning and/or after cloning. Restriction enzyme digestion is the preferred method for such fragmentation - though other methods (e.g. shearing or sonication) will be obvious to those skilled in the art.

For the particular case wherein the DNA is prepared from a library of clones (either genomic clones or cDNA clones) - with inserts derived from the individual(s) and propagated in some appropriate host and cloning vector system and wherein restriction enzyme fragmentation is used prior to cloning, polymerase chain reaction amplification can be used to prepare the DNA for comparison in fragmented form. Priming sites within the vector sequence flanking the cloned restriction enzyme fragmented inserts may be usefully employed for one or more cycles of polymerase chain reaction amplification of the fragmented DNA of interest. The primers used for polymerase chain reaction amplification of the fragmented DNA of interest could again be used after the phenotype-determining fragment enrichment process to 'rescue' and clone the enriched fragments.

Within the scope of the current invention, the terms biotinylation and streptavidin capture are used both as an example and as the currently preferred embodiment for the invention. The streptavidin may be surface attached to inert particles (magnetic or otherwise) or to vessel walls (e.g. microtitre plate wells). The biotin may be introduced via a deoxynucleotide triphosphate analogue using a polymerase; by using a biotin-conjugated primer and polymerase chain reaction amplification; chemically or photochemically. The use of biotin and streptavidin is not a limitation for the invention. The invention could equally be used with other high affinity capture systems well known to those skilled in the art (e.g. 'his tag' introduction and metal ion affinity capture).

Within the scope of the current invention, the term 'abnormal' - used with respect to the term 'normal' - is used without limitation in order

to denote a somatic cell (or somatic cells) with a discernable phenotypic characteristic (or characteristics) arising from the acquisition of a different somatic mutation (or set of somatic mutations) from that (or those) seen in the 'normal' counterpart. Cells will most usually be considered 'abnormal' with respect to their 'normal' counterparts through the acquisition of a different somatic mutation (or set of somatic mutations) leading to one or more of the following phenotypic characteristics: altered marker gene expression, altered genomic organisation, growth under certain selective culture conditions, immortalised growth in culture, unrestrained growth *in vivo* or *in vitro*, failure of normal apoptotic control mechanisms *in vivo* or *in vitro*, induction of neovascularisation, escape of cells across epithelium, migratory cell survival or metastasis.

Within the scope of the current invention, the term mismatch recognition protein is used without limitation to denote a protein of eukaryotic, prokaryotic or archaebacterial origin capable of the selective recognition of (and binding to) a DNA duplex that is not perfectly matched along its entire length. Recognition of (and binding to) will be preferably for bases that are not engaged in correct Watson and Crick pairing and for small deletions or insertions. Many such proteins are known to those skilled in the art. Prokaryotic and eukaryotic mutS homologues, phage T4 endonuclease VII, phage T7 endonuclease I and the plant enzyme CEL-1 are just some examples.

Inter-population perfectly matched duplex depletion

In the inter-population perfectly matched duplex depletion approach, we compare (in fragmented form) the pooled DNA of 'affected' individuals with the pooled DNA of 'unaffected' individuals (both from populations as outbred and otherwise similar to each other as possible). We are only interested in those regions where differences occur between 'affected' and 'unaffected' DNA molecules. For populations as above, the only prevalent sequence differences within the 'affected' pool (compared to

the 'unaffected' pool) should be somewhere within the gene(s) (using the term gene in its widest sense to include exons, introns and all associated upstream and downstream regulatory sequences) that actually determine(s) their common phenotype. This means that we are no longer tied into working with rare (and perhaps atypical) populations where there is high genetic homogeneity.

Pooling the DNA from entire phenotypically-defined populations massively reduces the amount of labour involved.

DNA sequence variation in populations

Nickerson *et al*, Nature Genetics, Vol 19, p233 (1998), sequenced 9.7 kb of the lipoprotein lipase gene from 71 individuals (24 African-Americans, 24 Europeans and 23 European-Americans). This gene is fairly typical (90 % intron and 10 % exon - total size 30 kb with 10 exons).

88 sequence variants were found (i.e. one per 110 bp on average). Most variations were found in non-coding sequence. 90 % of these were SNPs (60 % of which were transitions and 40 % were transversions). All of the SNPs were biallelic. 10 % of the sequence variants were insertions or deletions at repeat sequences.

58 % of the sequence variants were present in all three ethnic populations. Half of these were found in heterozygous form and half in homozygous form.

Nucleotide diversity (defined as the expected number of nucleotide differences per site between a random pair of chromosomes drawn from the population) is 1/500 for DNA in general and 1/2,000 for coding sequence DNA. This means that, on average, any two DNA fragments annealed together from such a population will contain a mismatch every 500 bases.

DNA sequence variants are therefore very common. They are not, however, totally random - the variants that occur every 500 bases or so are limited; they are generally biallelic at just that single base. It is

this fact that the inter-population perfectly matched duplex depletion approach selectively exploits.

Statistical analysis of the inter-population perfectly matched duplex depletion fragmentation process

If the length of the DNA fragments is F bases and the average length between sequence differences between any two DNA molecules is 500 bases, the probability that a hybrid duplex between any two random DNA fragments will contain no mismatches is given by

$$Pr(0) = e^{-(F/500)}$$

and the probability that a hybrid duplex between any two random DNA molecules will contain any mismatches is given by

$$Pr(\geq 1) = \{1-(e^{-(F/500)})\}$$

Example values for Pr(0) and Pr(≥1) for different values of F are given below

F	Pr(0)	Pr(≥1)
10	0.98	0.02
20	0.96	0.04
50	0.90	0.10
100	0.82	0.18
200	0.67	0.33
300	0.55	0.45
400	0.45	0.55
500	0.37	0.63
600	0.30	0.70
700	0.25	0.75
800	0.20	0.80
900	0.17	0.83
1000	0.14	0.86

Example average restriction fragment sizes for DNA digestion with six 6 bp cutters and up to four 4 bp cutters

Each 6 bp cutter will cut DNA every 4,096 bp on average and each 4 bp cutter will cut DNA every 256 bp on average.

For a given set of A 6 bp cutters and B 4 bp cutters (with no duplication of restriction enzyme cutting sequence), the average fragment length F will be

$$\{(A/4{,}096)+(B/256)\}^{-1}$$

Example values for F as A and B are varied are given in the following table

A	B	average size (F)
6	0	683
6	1	186
6	2	108
6	3	76
6	4	59

We should note that the above situation assumes that none of the 4 bp cutter recognition sites lie within any of the 6 bp cutter recognition sites. If, for example, we have a 4 bp cutter recognition site nested within a 6 bp cutter recognition site (e.g. from the use of *MboI* and *BamHI* in the fragmentation), then we should reduce the value of A from 6 to 5 in the above.

In general, if we have a given set of A enzymes that cut DNA every a bp on average, B enzymes that cut DNA every b bp on average, ... , Z enzymes that cut DNA every z bp on average (with no duplication of restriction enzyme cutting sequence), the average fragment length F will be

$$\{(A/a)+(B/b)+ \ldots + (Z/z)\}^{-1}$$

Example sets of 6 bp cutters and 4 bp cutters for fragmentation

Example sets of 6 bp cutters and 4 bp cutters that contain panels of six 6 bp cutters that are compatible with terminal restriction site profiling array (TRSPA) analysis (see below) are given in the following:

Example enzyme set 1

For restriction in M buffer + BSA

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
1	*MboI*	GATC	100	37	25	NEB	65/20
2	*HaeIII*	GGCC	100	37	50	NEB	80/20
3	*MseI*	TTAA	100	37	20	NEB	65/20

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
1	*BamHI*	GGATCC	100	37	100	NEB	80 / 20
2	*BsrGI*	TGTACA	100	37	10	NEB	80 / 20
3	*HindIII*	AAGCTT	100	37	>40	APB	65 / 20
4	*NcoI*	CCATGG	100	37	50	NEB	65 / 20
5	*SpeI*	ACTAGT	100	37	50	NEB	65 / 20
6	*AflII*	CTTAAG	100	37	10	APB	60 / 15

Example enzyme set 2

For restriction in M buffer + BSA

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
1	*MboI*	GATC	100	37	25	NEB	65/20
2	*HaeIII*	GGCC	100	37	50	NEB	80/20
3	*MseI*	TTAA	100	37	20	NEB	65/20

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
1	*EcoRI*	GAATTC	100	37	>40	APB	65 / 20
2	*BspHI*	TCATGA	100	37	10	APB	65 / 20
3	*BglII*	AGATCT	75	37	>40	APB	No
4	*XbaI*	TCTAGA	100	37	100	NEB	65 / 20
5	*Acc65I*	GGTACC	75	37	10	NEB	65 / 20
6	*ApaLI*	GTGCAC	100	37	10	NEB	No

Aspect (I)

In one aspect the invention provides a method of providing a mixture of DNA fragments enriched in fragments that are characteristic of a phenotype of interest, by providing affected DNA in fragmented form and providing unaffected DNA in fragmented form, which method comprises:

a) mixing the fragments of the affected DNA and the fragments of the unaffected DNA under hybridising conditions;

b) recovering a mixture of hybrids that contain mismatches;

c) recovering fragments of the affected DNA from the mixture of hybrids that contain mismatches;

and optionally repeating steps a), b) and c) one or more times.

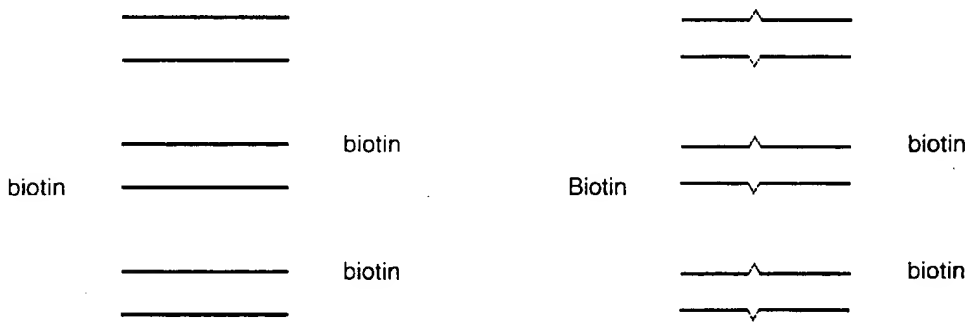
Inter-population mismatch-containing duplex selection

'Affected' versus 'unaffected' (i.e. inter-population) mismatch-containing duplex selection can be achieved by attaching a mismatch-binding protein to a solid support (or using the mismatch-binding protein in solution followed by subsequent solid-phase capture), taking fragmented and denatured 'affected' DNA and hybridising this to an excess of fragmented, denatured and biotinylated 'unaffected' DNA with ensuing capture of mismatch-containing duplex molecules. Releasing the mismatch-containing duplex molecules without denaturation, streptavidin

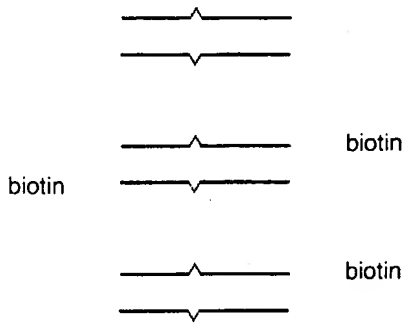
capture and then release of the non-biotinylated strands will give only the desired species as shown below.

Method

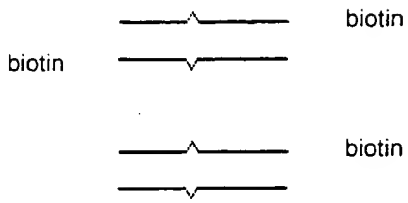
Fragment the 'affected' DNA. Fragment and derivatise the 'unaffected' DNA with biotin. Only DNA from this population will be streptavidin captured. Melt and anneal to give



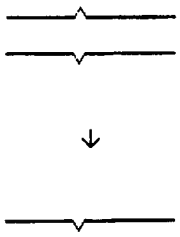
Mismatch-binding protein select. Capture only the mismatch-containing duplexes. Release without denaturation to give



Streptavidin capture to give



Release the non-biotinylated strands to give



Repeat as necessary.

Repetition of the above sequence of reactions will lead to inter-population perfectly matched duplex depletion.

What will be purified ?

Inter-population mismatch-containing duplex selection as above ensures that all of the various phenotype-determining fragments (unique to the 'affected' population) are captured for subsequent analysis - but it also causes the co-purification of very many SNP-containing ('noise') fragments.

We now need to consider the fate of the various types of fragment (i.e. those that determine the phenotype and those that do not) as we carry out inter-population perfectly matched duplex depletion cycles as above.

Recovery of 'affected' DNA molecules after streptavidin capture

For a particular fragment, if we have X molecules of 'affected' DNA and Y molecules of 'unaffected' biotinylated DNA, after complete hybridisation, there will be a ratio of {Y/(X+Y)} streptavidin-capturable molecules to {X/(X+Y)} streptavidin-non-capturable molecules. We can thus manipulate the yield of streptavidin-capturable hybrids by varying X and Y.

Example recovery and loss figures for various X and Y are shown in the following table

ratio (Y/X)	recovery = {Y/(X+Y)}	Loss = {X/(X+Y)}
1	50 %	50 %
2	67 %	33 %
3	75 %	25 %
4	80 %	20 %
9	90 %	10 %
19	95 %	5 %
99	99%	1 %
999	99.9 %	0.1 %

After n cycles, the recovery for a phenotype-determining fragment will be given by $\{Y/(X+Y)\}^n$

Loss of general SNP-containing ('noise') fragments during inter-population perfectly matched duplex depletion cycles

If we anneal fragmented DNA molecules together and capture only the mismatch-containing duplexes, then n repetitions of such a process will reduce the original number of fragments to the following fraction

$$\{1-(e^{-(F/500)})\}^n \cdot \{Y/(X+Y)\}^n$$

The enrichment for phenotype-determining fragments over SNP-containing ('noise') fragments during inter-population perfectly matched duplex depletion cycles will therefore be given by

$$\{1-(e^{-(F/500)})\}^{-n}$$

Example figures for enrichment are given below with F = 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 and 1,000 and n = 1, 2, 3, 4, 5, 6 ,7, 8, 9 and 10.

	n =									
F	1	2	3	4	5	6	7	8	9	10
50	11	110	1,160	12,194	128,135	1,346,489	1.E+07	1.E+08	2.E+09	2.E+10
100	6	30	168	926	5,110	28,187	155,500	857,841	4,732,415	3.E+07
200	3	9	28	85	257	779	2,362	7,166	21,735	65,929
300	2	5	11	24	53	119	263	582	1,291	2,860
400	2	3	6	11	20	36	65	118	215	390
500	2	3	4	6	10	16	25	39	62	98
600	1	2	3	4	6	9	12	18	25	36
700	1	2	2	3	4	5	7	10	13	17
800	1	2	2	2	3	4	5	6	8	10
900	1	1	2	2	2	3	4	4	5	6
1,000	1	1	2	2	2	2	3	3	4	4

Loss of specific SNP-containing ('noise') fragments during inter-population perfectly matched duplex depletion cycles

The non-polymorphic and SNP-containing ('noise') fragments will be depleted as described above.

Not all fragments will, however, be depleted at the same rate. An individual SNP-containing ('noise') fragment will be depleted with every cycle of inter-population perfectly matched duplex depletion as outlined below.

Let us assume that there are two alleles for the polymorphism within a particular fragment. Let these be called P and Q. Let p be the

fraction of the P allele in the (outbred) 'affected' and 'unaffected' populations and let q be the fraction of the Q allele. Let (p+q)=1, so that q=(1-p).

After denaturation and annealing, the four possible events are PP, PQ, QP and QQ. PP and QQ will form perfectly matched duplexes and will therefore be lost - whereas PQ and QP will form mismatch-containing duplexes and will consequently be recovered.

After one cycle, the fraction of recovered molecules will be

$$\{2pq/(p^2+2pq+q^2)\}$$

$$= \{2pq/((p+q)^2)\}$$

$$= 2pq$$

$$= 2p(1-p)$$

Hence if we start out with M molecules of DNA from the population, there will be 2Mpq molecules remaining after the first round of inter-population mismatch-containing duplex selection. Let us denote the number of molecules entering the second round of inter-population mismatch-containing duplex selection by M', where

$$M' = 2Mpq$$

and the fraction of lost molecules will be

$$\{(p^2+q^2)/(p^2+2pq+q^2)\}$$

$$= \{(p^2+q^2)/((p+q)^2)\}$$

$$= p^2+q^2$$

$$= p^2+\{1-2p+p^2\}$$

$$= 1-2p+2p^2$$

$$= 1-2p(1-p)$$

$$=1-2pq$$

The fractional loss of P-allelic molecules will be p^2 and the fractional loss of Q-allelic molecules will be q^2 (= $(1-p)^2$).

If we start out with M molecules of DNA from the population, there will be Mp of the P-allelic molecules and Mq = M(1-p) of the Q-allelic molecules before inter-population mismatch-containing duplex selection. After inter-population mismatch-containing duplex selection, there will therefore be $Mp-Mp^2 = Mp(1-p) = Mpq$ of the P-allelic molecules and $Mq-Mq^2 = Mq(1-q) = Mqp$ of the Q-allelic molecules. In other words, after the first round of mismatch-containing duplex selection, there will be the same number of P-allelic molecules as Q-allelic molecules.

We can define new allelic frequencies p' and q' as follows

$$p' = q' = 0.5$$

If we now perform a second round of inter-population mismatch-containing duplex selection, we start out with M' molecules of DNA from the first round. There will be M'p' of the P-allelic molecules and M'q' of the Q-allelic molecules before inter-population mismatch-containing duplex selection.

After inter-population mismatch-containing duplex selection, there will therefore be $M'p'-Mpp' = M'p'(1-p) = M'p'q$ of the P-allelic

molecules and M'q'-Mqq' = M'q'(1-q) = M'q'p of the Q-allelic molecules.

The total number of molecules (M") will be

$$M'p'q + M'q'p$$

$$= M'(p'q-q'p)$$

but $p' = q' = 0.5$

hence $M" = 0.5*M'(p+q)$

but $(p+q) = 1$

so $M" = (M'/2)$

In other words, after the second round of inter-population mismatch-containing duplex selection, there will again be the same number of P-allelic molecules as Q-allelic molecules. Thus the new allelic frequencies p" and q" remain as previously p" = q" = 0.5.

A pattern now emerges. After the first cycle of inter-population mismatch-containing duplex selection, the allelic frequencies are equalled and the number of molecules is reduced to 2Mpq (Mpq of each allelic molecule). Thereafter, every cycle halves the number of molecules and keeps both alleles at the same frequency.

Consequently, after n cycles, the number of P-allelic molecules will be

$$Mpq(0.5)^{n-1}$$

and the number of Q allelic molecules will be

$$Mqp(0.5)^{n-1}$$

both are, of course, equal.

If we now take the capture yield (see above) into consideration, the SNP-containing ('noise') fragment yield will be given by

$$2Mpq(0.5)^{n-1}\cdot\{Y/(X+Y)\}^{n}$$

where both allelic variants are deemed to be captured 'noise'.

Polymorphisms that interfere with the pattern of restriction digestion

For both the loss of general and specific SNP-containing ('noise') fragments during inter-population perfectly matched duplex depletion cycles (described above) and where the SNP interferes with the pattern of restriction digestion, if the mismatch-binding protein also binds to duplex molecules with unequal lengths (e.g. from inter-population annealing around a site of restriction site polymorphism), then the above analysis still holds (with perfectly matched duplex being replaced by equal length duplex and mismatch-containing duplex being replaced by unequal partner-length duplex).

In the rare cases where a restriction site is lost due to a sequence change that actually determines the phenotype of interest, a double-length fragment will be obtained. This will give rise to a double terminal restriction site profiling array (TRSPA) signature (see below). Multiple isolates of the particular double signature will be indicative of an association between the fragment and the phenotype of interest.

Further 'kinetic' enrichment to enhance the selective removal of SNP-containing 'noise' from the pool of phenotype determining fragments

After multiple cycles of enrichment by the above procedure, the enriched DNA pool should contain many copies of all phenotype-determining fragments but also low numbers of copies of many different phenotype non-determining fragments. The total number of 'noise' fragments may exceed the number of phenotype determining fragments, despite the number of each individual 'noise' species being very small. The 'noise' fragments would therefore increase the number of probes required for TRSPA analysis before a pattern emerges. To largely eliminate this problem, a further kinetic enrichment procedure is used. Either one or both of strategies A and B below can be employed to achieve 'kinetic' enrichment.

Strategy A – Subtraction of the enriched DNA from inter-population mismatch containing duplex depletion

The enriched fragment pool from inter-population mismatch containing duplex depletion is rapidly self-hybridised - enabling the common phenotype-determining fragments to form perfectly matched duplexes with greater efficiency than the rare 'noise' fragments. Selection for perfectly matched duplexes then yields a selectively further enriched pool of fragments. Multiple cycles of subtraction could be carried out if necessary.

Strategy B – Hybridisation of the enriched DNA from inter-population mismatch containing duplex depletion against the 'affected' DNA pool

The enriched fragment pool from inter-population mismatch containing duplex depletion is then hybridised to an excess of biotinylated DNA from the 'affected' pool. This allows the common phenotype-determining fragments to form perfectly matched duplexes with greater efficiency than the rare 'noise' fragments. Selection for perfectly matched

duplexes followed by streptavidin capture and denaturation to release the non-biotinylated strands then yields a further enriched pool of fragments. Multiple such 'affected' pool back-hybridisations could be carried out if necessary.

Extension of the invention to the case of single phenotypically 'affected' individuals within populations where the distinction between 'affected' and 'unaffected' is clear

The above has described inter-population perfectly matched duplex depletion between non-biotinylated DNA fragments from an 'affected' population and biotinylated DNA fragments from an 'unaffected' population. Provided the 'unaffected' population is sufficiently complex that it contains all the non-phenotype-determining sequence variants found in a single 'affected' individual, then inter-population perfectly matched duplex depletion should be possible for single phenotypically 'affected' individuals against an 'unaffected' population where the distinction between 'affected' and 'unaffected' is clear. The latter proviso is needed in order to ensure that a small number of misdiagnosed 'affected' individuals in the 'unaffected' population do not cause the removal of phenotype-determining fragments during inter-population perfectly matched duplex molecular depletion.

Extension of the invention to the case of disease gene identification in cases where novel phenotype-determining mutations arise spontaneously within a family

Except for a small number of sequence changes, each of us contains DNA sequence derived from our parents - our individuality resulting from precisely which parental alleles we receive. If one of the above small number of sequence changes results in a change in phenotype, then we can use inter-population perfectly matched duplex depletion to enrich for fragments encoding this change in phenotype.

If we take 'unaffected total ancestral' cells (by which we mean cells derived from a complete set of 'unaffected' ancestors - e.g. both parents, or mother plus two paternal grandparents, or father plus two maternal grandparents, or two maternal grandparents and two paternal grandparents etc.) as the source of our biotinylated fragments and cells from an 'affected' descendent as the source of our non-biotinylated fragments, any fragments that have acquired phenotype-determining sequence changes between 'unaffected' ancestral generations and the 'affected' descendent generation will be unable to form perfectly matched duplexes with the biotinylated 'unaffected total ancestral' fragments. Successive cycles of such inter-population perfectly matched duplex depletion will thus lead to the enrichment of fragments carrying all such sequence - the degree of enrichment per cycle being as described below.

Statistical analysis

Let us assume that equal numbers of fragments are used from each of the 'unaffected' ancestors. Let the number of such ancestors be A.

{1/A} of the annealings in the inter-population perfectly matched duplex depletion will be self-against-ancestral-transmitted alleles - statistically equivalent to self-against-self inter-population perfectly matched duplex depletion (see below).

{(A-1)/A} of the annealings in the inter-population perfectly matched duplex depletion will be self-against-ancestral-nontransmitted alleles - statistically equivalent to inter-population perfectly matched duplex depletion between unrelated individuals.

A	{1/A}	{(A-1)/A}
2	1/2	½
3	1/3	2/3
4	1/4	¾

From the data of Nickerson *et al*, a DNA sequence variation between unrelated individuals should occur every 500 base pairs (it is this figure that we use for inter-population perfectly matched duplex depletion with self-against-ancestral-nontransmitted alleles and inter-population perfectly matched duplex depletion between unrelated individuals). In addition, a given individual should be heterozygous once every 573 base pairs (this figure is used for inter-population perfectly matched duplex depletion with self-against-ancestral-transmitted alleles and self-against-self inter-population perfectly matched duplex molecular depletion). Inter-population perfectly matched duplex depletion against the transmitted alleles and the nontransmitted alleles will now be considered separately.

Self-against-ancestral-transmitted alleles inter-population perfectly matched duplex depletion

If we anneal the two complementary strands for a fragment using DNA from an 'affected' descendent and DNA containing the transmitted alleles of 'unaffected' ancestors, then $\{1-e^{(-F/573)}\}$ of fragments will contain one or more site of heterozygosity. {1/A} of the annealings will be of this type. For such annealings, where a site of heterozygosity is present, the probability of obtaining a mismatch-containing duplex between a biotinylated fragment and a non-biotinylated fragment containing the site of heterozygosity is 0.5.

Self-against-ancestral-nontransmitted alleles inter-population perfectly matched duplex depletion

If we anneal the two complementary strands for a fragment using DNA from an 'affected' descendent and DNA containing the nontransmitted alleles of 'unaffected' ancestors, then $\{1-e^{(-F/500)}\}$ of fragments will contain one or more site of DNA sequence variation. {(A-1)/A} annealings will be of this type.

Phenotype-determining fragment enrichment

The fraction of fragments carried through the first cycle of inter-population perfectly matched duplex depletion will therefore be

$$([0.5\cdot\{1/A\}\cdot\{1-e^{(-F/573)}\}]+[\{(A-1)/A\}\cdot\{1-e^{(-F/500)}\}])\cdot\{Y/(X+Y)\}$$

where {Y/(X+Y)} represents the yield of streptavidin-capturable molecules.

Hence n repetitions of such a process will reduce the original number of fragments to the following fraction

$$([0.5\cdot\{1/A\}\cdot\{1-e^{(-F/573)}\}]+[\{(A-1)/A\}\cdot\{1-e^{(-F/500)}\}])^{n}\cdot\{Y/(X+Y)\}^{n}$$

The enrichment for phenotype-determining fragments over SNP-containing ('noise') fragments during inter-population perfectly matched duplex depletion cycles will therefore be given by

$$([0.5\cdot\{1/A\}\cdot\{1-e^{(-F/573)}\}]+[\{(A-1)/A\}\cdot\{1-e^{(-F/500)}\}])^{-n}$$

Example figures for enrichment are given for A = 2, 3 and 4 below with F = 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 and 1,000 and n = 1, 2, 3, 4 , 5 ,6 ,7, 8, 9 and 10.

For A=2

	n=									
F	1	2	3	4	5	6	7	8	9	10
50	15	213	3,115	45,494	664,427	9,703,694	1.E+08	2.E+09	3.E+10	4.E+11
100	8	59	448	3,430	26,250	200,887	1,537,367	1.E+07	9.E+07	7.E+08
200	4	18	74	309	1,296	5,434	22,782	95,524	400,520	1,679,340
300	3	9	28	87	265	811	2,475	7,558	23,078	70,468
400	2	6	16	39	97	241	600	1,497	3,734	9,313
500	2	5	10	22	48	103	224	485	1,051	2,277
600	2	4	7	15	29	56	109	213	416	813
700	2	3	6	11	19	35	63	114	207	374
800	2	3	5	8	14	24	42	71	121	205
900	2	3	4	7	11	18	30	49	79	128
1,000	2	2	4	6	9	15	23	36	57	89

For A=3

	n=									
F	1	2	3	4	5	6	7	8	9	10
50	13	167	2,154	27,823	359,340	4,640,941	6.E+07	8.E+08	1.E+10	1.E+11
100	7	46	311	2,104	14,251	96,516	653,682	4,427,241	3.E+07	2.E+08
200	4	14	51	191	709	2,633	9,786	36,365	135,135	502,173
300	3	7	20	54	146	396	1,073	2,909	7,882	21,361
400	2	5	11	24	53	119	263	582	1,290	2,859
500	2	4	7	14	27	51	99	190	367	707
600	2	3	5	9	16	28	48	84	147	255
700	2	3	4	7	11	18	28	46	74	119
800	2	2	4	5	8	12	19	28	43	66
900	1	2	3	4	6	9	14	20	29	41
1,000	1	2	3	4	5	8	11	15	21	29

For A=4

	n=									
F	1	2	3	4	5	6	7	8	9	10
50	12	149	1,826	22,316	272,759	3,333,771	4.E+07	5.E+08	6.E+09	7.E+10
100	6	41	264	1,690	10,834	69,465	445,373	2,855,512	2.E+07	1.E+08
200	4	12	44	154	540	1,902	6,696	23,570	82,967	292,045
300	3	7	17	44	112	287	737	1,894	4,864	12,494
400	2	4	9	20	41	86	181	381	800	1,681
500	2	3	6	11	20	37	68	125	229	418
600	2	3	5	7	12	20	34	56	92	152
700	2	2	4	5	8	13	20	30	46	71
800	1	2	3	4	6	9	13	19	27	39
900	1	2	3	4	5	7	9	13	18	25
1,000	1	2	2	3	4	6	7	10	13	17

Extension of the invention to the case of fully comprehensive 'abnormal' cell mutational profiling within an individual

If we now take 'normal' cells as the source of our biotinylated fragments and 'abnormal' cells from the same individual as the source of our non-biotinylated fragments, any fragments that have acquired sequence changes on the way to becoming 'abnormal' will be unable to form perfectly matched duplexes with the biotinylated 'normal' fragments. Successive cycles of such inter-population perfectly matched duplex depletion will thus lead to the enrichment of fragments carrying all those sequence differences between the 'normal' cells and the 'abnormal' cells - the degree of enrichment per cycle being as described below.

From the data of Nickerson *et al*, a given individual should be heterozygous about once every 573 base pairs.

If we anneal the two complementary strands for a fragment using DNA from 'abnormal' cells and DNA from 'normal' cells, then $\{1-e^{(-F/573)}\}$ of fragments will contain one or more site of heterozygosity.

For a heterozygous site, p and q are both 0.5. The probability of obtaining a perfectly matched duplex between a biotinylated fragment

and a non-biotinylated fragment containing the site is 0.5. Similarly, the probability of obtaining a mismatch-containing duplex between a biotinylated fragment and a non-biotinylated fragment containing the site is also 0.5.

The fraction of fragments carried through the first cycle of inter-population perfectly matched duplex depletion will therefore be

$$0.5 \cdot \{1-e^{(-F/573)}\} \cdot \{Y/(X+Y)\}$$

hence n repetitions of such a process will reduce the original number of fragments to the following fraction

$$[0.5 \cdot \{1-e^{(-F/573)}\}]^{n} \cdot \{Y/(X+Y)\}^{n}$$

The enrichment for phenotype-determining fragments over SNP-containing ('noise') fragments during inter-population perfectly matched duplex depletion cycles will therefore be given by

$$[0.5 \cdot \{1-e^{(-F/573)}\}]^{-n}$$

Example figures for enrichment are given below with F = 50, 100, 200, 300, 400, 500, 600, 700, 800, 900 and 1,000 and n = 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10.

	n=									
F	1	2	3	4	5	6	7	8	9	10
50	24	573	13.711	328.171	7.854.627	2.E+08	4.E+09	1.E+11	3.E+12	6.E+13
100	12	156	1.948	24.329	303,844	3.794,731	5.E+07	6.E+08	7.E+09	9.E+10
200	7	46	313	2.123	14.412	97,830	664.076	4.507.784	3.E+07	2.E+08
300	5	24	118	580	2.845	13.958	68.490	336.072	1.649,059	8,091,708
400	4	16	63	251	999	3.977	15.831	63.012	250.815	998,346
500	3	12	41	139	479	1.644	5.650	19.411	66,688	229,114
600	3	9	29	90	278	856	2.638	8.128	25,047	77,179
700	3	8	23	65	183	520	1.475	4,183	11,862	33,639
800	3	7	19	50	133	353	937	2,491	6,621	17,599
900	3	6	16	41	103	259	654	1.652	4,171	10,532
1,000	2	6	14	34	84	202	490	1,188	2,880	6,978

This approach will enrich for fragments containing *all* sequence differences within the 'abnormal' cells - no prior knowledge of the genes (oncogenes, tumour suppressor genes etc.) that may need to be investigated is required.

Having thus isolated a fragment (or fragments) determining differences between the 'affected' and 'unaffected' populations, we can now proceed to their analysis.

Aspect (II)

In another aspect the invention provides a method of making a set of arrays of fragments of DNA of interest, which method comprises:

a) selecting, from a set of n restriction endonuclease enzymes, a subset of r restriction endonuclease enzymes;

b) digesting genomic DNA with the subset of r enzymes;

c) ligating to the resulting fragments restriction-enzyme-cutting-site-specific adapters with unique polymerase chain reaction amplifiable sequences;

d) splitting the resulting fragments into r^2 aliquots;

e) amplifying each aliquot with two-restriction enzyme-specific

primers;

f) forming an array of the r^2 aliquots of the amplimers; and

g) repeating steps a) to f) using a different subset of r restriction endonuclease enzymes. The invention also includes sets of arrays obtained or obtainable by the method.

The n restriction endonuclease enzymes may be selected from 4-cutters and 5-cutters and 6-cutters, and a set may include enzymes from one or two or three of these categories. The value of n is preferably 3 to 10, for reasons discussed below. The value of r is less than n and is preferably 2 to 4, chosen with reference to the frequency with which the chosen enzymes cut nucleic acids, and ease of fragment amplification by PCR.

Terminal restriction site profiling arrays (TRSPAs)

If we use a total of n 6 bp cutter restriction enzymes within the total set of enzymes used for fragmentation, let us use subsets of r 6 bp cutter enzymes (taken from the total set n) to make (rxr) TRSPA test matrices as follows.

Version 1 - (TRSPA-1)

For each (rxr) TRSPA test matrix, digest DNA to completion with all r restriction enzymes. Ligate on restriction enzyme cutting-site-specific adaptors with unique polymerase chain reaction amplifiable tags. Split into r^2 aliquots and for each aliquot, amplify with biotinylated restriction $enzyme_j$ and non-biotinylated restriction $enzyme_k$ tag-specific primers and array the non-biotinylated strands for all values of j and k between 1 and r.

Example

Consider the following dsDNA

--[A]--1--[B]--2--[B]--3--[C]--4--[A]--5--[B]-- +

--[A]--1--[B]--2--[B]--3--[C]--4--[A]--5--[B]-- -

where the restriction enzyme cutting sites are denoted A, B and C and the fragments after restriction digestion are denoted 1, 2, 3, 4 and 5. + and - denote the sense of the strands.

Cut to completion with A, B and C to give

A--1--B B--2--B B--3--C C--4--A A--5--B +

A--1--B B--2--B B--3--C C--4--A A--5--B -

Polymerase chain reaction amplify according to the following primer matrix and streptavidin capture to give

	A	B	C
bio-A		bio-A--1--B + A--1--B - bio-A--5--B + A--5--B -	C--4--A + C--4--A-bio -
bio-B	A--1--B + A--1--B-bio - A--5--B + A--5--B-bio -	bio-B--2--B + B--2--B - B--2--B + B--2--B-bio -	bio-B--3--C + B--3--C -
bio-C	bio-C--4--A + C--4--A -	B--3--C + B--3--C-bio -	

Keep only the non-biotinylated strands to give

	A	B	C
bio-A		A--1--B - A--5--B -	C--4--A +
bio-B	A--1--B + A--5--B +	B--2--B - B--2--B +	B--3--C -
bio-C	C--4--A -	B--3--C +	

In simplified form

	A	B	C
bio-A		1-. 5-	4+
bio-B	1+. 5+	2-, 2+	3-
bio-C	4-	3+	

Repeat for all ${}^{n}C_{r}$ combinations of restriction enzymes to generate the TRSPA.

TRSPA-1 hybridisation patterns

The two types of TRSPA-1 hybridisation pattern we should expect using a probe resulting from inter-population perfectly matched duplex depletion are

1. Hybridisation to an off-diagonal element (e.g. row x, column y - where x and y are different) and its complementary element reflected across the diagonal (i.e. row y, column x) and
2. Hybridisation to an on-diagonal element (e.g. the element at row z, column z).

TRSPA-1 analysis - a worked example for n=3 and r=2

Let us take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

where

A, B and C denote restriction enzyme cutting sites. 1, 2, 3, 4, 5, 6, 7, 8 and 9 denote the restriction fragments after digestion.

The TRSPA-1 test matrices

There are ${}^{3}C_{2}=3$ TRSPA-1 test matrices - AB, BC and AC.

The test matrix hybridisation patterns

There are {(2·(2+1))/2}=3 possible hybridisation patterns for each of the test matrices

The combinatorial diversity

If there are three possible TRSPA-1 test matrix hybridisation patterns and three possible test matrices, then there will be 3^3=27 possible TRSPA-1 signatures.

Fragment 4 analysis - the AB TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and B to give

-C-1-B

B-2-B

B-3-A

A-4-A

A-5-C-6-C-7-B

B-8-A

A-9-A-

The fragment complementary to fragment 4 will be

A-4-A

Polymerase chain reaction amplify and hybridise with just fragment 4 to give

AB matrix	A	B
bio-A		
bio-B		

Fragment 4 analysis - the BC TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with B and C to give

-C-1-B

B-2-B

B-3-A-4-A-5-C

C-6-C

C-7-B

B-8-A-9-A-

The fragment complementary to fragment 4 will be

B-3-A-4-A-5-C

Polymerase chain reaction amplify and hybridise with just fragment 4 to give

BC matrix	B	C
bio-B		
bio-C		

Fragment 4 analysis - the AC TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and C to give

-C

C-1-B-2-B-3-A

A-4-A

A-5-C

C-6-C

C-7-B-8-A

A-9-A

A-

The fragment complementary to fragment 4 will be

A-4-A

Polymerase chain reaction amplify and hybridise with just fragment 4 to give

AC matrix	A	C
bio-A		
bio-C		

Fragment 5 analysis - the AB TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and B to give

-C-1-B

B-2-B

B-3-A

A-4-A

A-5-C-6-C-7-B

B-8-A

A-9-A-

The fragment complementary to fragment 5 will be

A-5-C-6-C-7-B

Polymerase chain reaction amplify and hybridise with just fragment 5 to give

AB matrix	A	B
bio-A		
bio-B		

Fragment 5 analysis - the BC TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with B and C to give

-C-1-B

B-2-B

B-3-A-4-A-5-C

C-6-C

C-7-B

B-8-A-9-A-

The fragment complementary to fragment 5 will be

B-3-A-4-A-5-C

Polymerase chain reaction amplify and hybridise with just fragment 5 to give

BC matrix	B	C
bio-B		
bio-C		

Fragment 5 analysis - the AC TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and C to give

-C

C-1-B-2-B-3-A

A-4-A

A-5-C

C-6-C

C-7-B-8-A

A-9-A

A-

The fragment complementary to fragment 5 will be

A-5-C

Polymerase chain reaction amplify and hybridise with just fragment 5 to give

AC matrix	A	C
bio-A		
bio-C		

Fragment 6 analysis - the AB TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and B to give

-C-1-B

B-2-B

B-3-A

A-4-A

A-5-C-6-C-7-B

B-8-A

A-9-A-

The fragment complementary to fragment 6 will be

A-5-C-6-C-7-B

Polymerase chain reaction amplify and hybridise with just fragment 6 to give

AB matrix	A	B
bio-A		
bio-B		

Fragment 6 analysis - the BC TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with B and C to give

-C-1-B

B-2-B

B-3-A-4-A-5-C

C-6-C

C-7-B

B-8-A-9-A-

The fragment complementary to fragment 6 will be

C-6-C

Polymerase chain reaction amplify and hybridise with just fragment 6 to give

BC matrix	B	C
bio-B		
bio-C		

Fragment 6 analysis - the AC TRSPA-1 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and C to give

-C

C-1-B-2-B-3-A

A-4-A

A-5-C

C-6-C

C-7-B-8-A

A-9-A

A-

The fragment complementary to fragment 6 will be

C-6-C

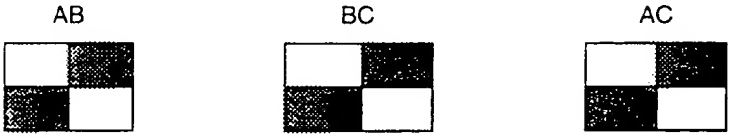
Polymerase chain reaction amplify and hybridise with just fragment 6 to give

AC matrix	A	C
bio-A		
bio-C		

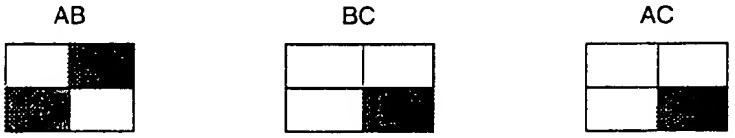
Overall results

Fragment 4 TRSPA-1 analysis

Fragment 5 TRSPA-1 analysis



Fragment 6 TRSPA-1 analysis



Version 2 - (TRSPA-2)

For each (rxr) TRSPA test matrix, digest DNA to completion with all r restriction enzymes. Ligate on restriction enzyme cutting-site-specific adaptors with unique polymerase chain reaction amplifiable tags. Split into r^2 aliquots and for each aliquot, amplify with non-biotinylated restriction enzyme$_j$ and non-biotinylated restriction enzyme$_k$ tag-specific primers and array the denatured strands for all values of j and k between 1 and r.

Example

Consider the following dsDNA

--[A]--1--[B]--2--[B]--3--[C]--4--[A]--5--[B]-- +
--[A]--1--[B]--2--[B]--3--[C]--4--[A]--5--[B]-- -

where the restriction enzyme cutting sites are denoted A, B and C and the fragments after restriction digestion are denoted 1, 2, 3, 4 and 5. + and - denote the sense of the strands.

Cut to completion with A, B and C to give

A--1--B B--2--B B--3--C C--4--A A--5--B +
A--1--B B--2--B B--3--C C--4--A A--5--B -

Polymerase chain reaction amplify according to the following primer matrix to give

	A	B	C
A		A--1--B + A--1--B - B--2--B + B--2--B - A--5--B + A--5--B -	C--4--A + C--4--A -
B	A--1--B + A--1--B - B--2--B + B--2--B - A--5--B + A--5--B -	B--2--B + B--2--B -	B--2--B + B--2--B - B--3--C + B--3--C -
C	C--4--A + C--4--A -	B--2--B + B--2--B - B--3--C + B--3--C -	

In simplified form

	A	B	C
A		1, 2, 5	4
B	1, 2, 5	2	2, 3
C	4	2, 3	

Repeat for all ${}^{n}C_{r}$ combinations of restriction enzymes to generate the TRSPA.

TSPSA-2 hybridisation patterns

The two types of TRSPA-2 hybridisation pattern we should expect using a probe resulting from inter-population perfectly matched duplex depletion are

1. Hybridisation to an off-diagonal element (e.g. row x, column y - where x and y are different) and its complementary element reflected across the diagonal (i.e. row y, column x) and
2. Hybridisation to a whole row and column intersecting at an on-diagonal element (e.g. all of row z and all of column z).

TRSPA-2 analysis - a worked example for n=3 and r=2

Let us take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

where

A, B and C denote restriction enzyme cutting sites. 1, 2, 3, 4, 5, 6, 7, 8 and 9 denote the restriction fragments after digestion.

The TRSPA-2 test matrices

There are ${}^{3}C_{2}=3$ TRSPA-2 test matrices - AB, BC and AC.

The test matrix hybridisation patterns

There are $\{(2\cdot(2+1))/2\}=3$ possible hybridisation patterns for each of the test matrices



The combinatorial diversity

If there are three possible TRSPA-2 test matrix hybridisation

patterns and three possible test matrices, then there will be 3^3=27 possible TRSPA-2 signatures.

Fragment 4 analysis - the AB TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and B to give

-C-1-B

B-2-B

B-3-A

A-4-A

A-5-C-6-C-7-B

B-8-A

A-9-A-

The fragment complementary to fragment 4 will be

A-4-A

Polymerase chain reaction amplify and hybridise with just fragment 4 to give

AB matrix	A	B
A		
B		

Fragment 4 analysis - the BC TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with B and C to give

-C-1-B

B-2-B

B-3-A-4-A-5-C

C-6-C

C-7-B

B-8-A-9-A-

The fragment complementary to fragment 4 will be

B-3-A-4-A-5-C

Polymerase chain reaction amplify and hybridise with just fragment 4 to give

BC matrix	B	C
B		
C		

Fragment 4 analysis - the AC TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and C to give

-C

C-1-B-2-B-3-A

A-4-A

A-5-C

C-6-C

C-7-B-8-A

A-9-A

A-

The fragment complementary to fragment 4 will be

A-4-A

Polymerase chain reaction amplify and hybridise with just fragment 4 to give

AC matrix	A	C
A		
C		

Fragment 5 analysis - the AB TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and B to give

-C-1-B

B-2-B

B-3-A

A-4-A

A-5-C-6-C-7-B

B-8-A

A-9-A-

The fragment complementary to fragment 5 will be

A-5-C-6-C-7-B

Polymerase chain reaction amplify and hybridise with just fragment 5 to give

AB matrix	A	B
A		
B		

Fragment 5 analysis - the BC TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with B and C to give

-C-1-B

B-2-B

B-3-A-4-A-5-C

C-6-C

C-7-B

B-8-A-9-A-

The fragment complementary to fragment 5 will be

B-3-A-4-A-5-C

Polymerase chain reaction amplify and hybridise with just fragment 5 to give

BC matrix	B	C
B		
C		

Fragment 5 analysis - the AC TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and C to give

-C
C-1-B-2-B-3-A
A-4-A
A-5-C
C-6-C
C-7-B-8-A
A-9-A
A-

The fragment complementary to fragment 5 will be

A-5-C

Polymerase chain reaction amplify and hybridise with just fragment 5 to give

AC matrix	A	C
A		
C		

Fragment 6 analysis - the AB TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and B to give

-C-1-B

B-2-B

B-3-A

A-4-A

A-5-C-6-C-7-B

B-8-A

A-9-A-

The fragment complementary to fragment 6 will be

A-5-C-6-C-7-B

Polymerase chain reaction amplify and hybridise with just fragment 6 to give

AB matrix	A	B
A		
B		

Fragment 6 analysis - the BC TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with B and C to give

-C-1-B

B-2-B

B-3-A-4-A-5-C

C-6-C

C-7-B

B-8-A-9-A-

The fragment complementary to fragment 6 will be

C-6-C

Polymerase chain reaction amplify and hybridise with just fragment 6 to give

BC matrix	B	C
B		
C		

Fragment 6 analysis - the AC TRSPA-2 matrix

Take the dsDNA

-C-1-B-2-B-3-A-4-A-5-C-6-C-7-B-8-A-9-A-

Cut with A and C to give

-C

C-1-B-2-B-3-A

A-4-A

A-5-C

C-6-C

C-7-B-8-A

A-9-A

A-

The fragment complementary to fragment 6 will be

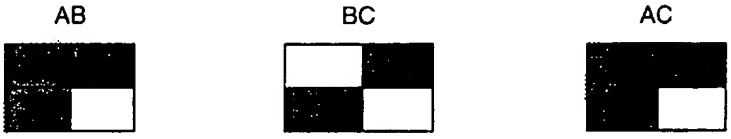
C-6-C

Polymerase chain reaction amplify and hybridise with just fragment 6 to give

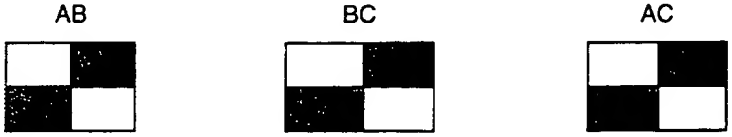
AC matrix	A	C
A		
C		

Overall results

Fragment 4 TRSPA-2 analysis



Fragment 5 TRSPA-2 analysis



Fragment 6 TRSPA-2 analysis

The number of test matrices

For a total of n enzymes used for fragmentation and a panel of r enzymes per (rxr) test matrix, there will be ${}^nC_r$ possible (rxr) test matrices, where

$${}^nC_r = (n!)/[(n-r)!\cdot r!]$$

${}^nC_r$ for various n and r is given in the following table

	n=							
r	3	4	5	6	7	8	9	10
1	3	4	5	6	7	8	9	10
2	3	6	10	15	21	28	36	45
3	1	4	10	20	35	56	84	120
4	-	1	5	15	35	70	126	210
5	-	-	1	6	21	56	126	252
6	-	-	-	1	7	28	84	210
7	-	-	-	-	1	8	36	120
8	-	-	-	-	-	1	9	45
9	-	-	-	-	-	-	1	10
10	-	-	-	-	-	-	-	1

The number of TRSPA spots arrayed

The total number of TRSPA spots is given by

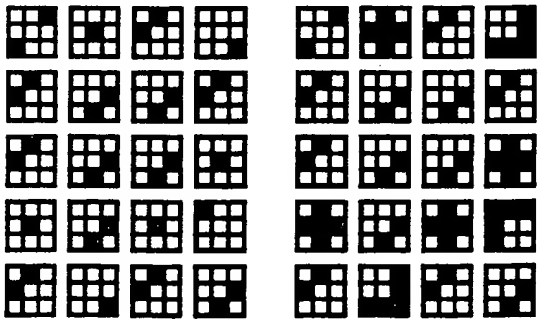
$$r^2\cdot\{{}^nC_r\} = r^2\cdot\{(n!)/[(n-r)!\cdot r!]\}$$

$r^2 \cdot \{^nC_r\}$ for various n and r is given in the following table

	n =							
R	3	4	5	6	7	8	9	10
1	3	4	5	6	7	8	9	10
2	12	24	40	60	84	112	144	180
3	9	36	90	180	315	504	756	1,080
4	-	16	80	240	560	1,120	2,016	3,360
5	-	-	25	150	525	1,400	3,150	6,300
6	-	-	-	36	252	1,008	3,024	7,560
7	-	-	-	-	49	392	1,764	5,880
8	-	-	-	-	-	64	576	2,880
9	-	-	-	-	-	-	81	810
10	-	-	-	-	-	-	-	100

Each test as above will give rise to a particular hybridisation 'signature'.

TRSPA-1 (for n=6 and r=3) TRSPA-2 (for n=6 and r=3)



There will be {1+2+...+r} = {r(r+1)/2} patterns per (rxr) test matrix. If there are ^nC_r (rxr) test matrices, the total possible number of signatures will be given by {r(r+1)/2} raised to the power ^nC_r. We require that the number of such different signatures is greatly in excess over the

number of fragments generated upon fragmentation.

If we choose n=6 and r=3, there will be 6 possible hybridisation patterns per test and 20 possible tests - giving a combinatorial diversity of $6^{20} = 3.7 \cdot 10^{15}$. The preferred scheme therefore employs six enzymes, tested in groups of three - giving 180 spots per experiment. In order to avoid restriction fragment length polymorphism problems, a duplicate analysis could be performed with a different set of enzymes sharing no cutting sites in common with the first set.

Example restriction enzyme sets for the preparation of test matrices for TRSPAs

Criteria for enzyme choice

In order to make TRSPAs, the selection of suitable enzymes is an important factor. Ideally, two sets of different enzymes are required to eliminate the small possibility that a phenotype-determining polymorphism might fall within a chosen restriction site and therefore compromise the specificity of the resulting signature. The selection of enzymes can be based upon a number of criteria

a) The enzymes should be 6 bp cutters.

b) Cleavage by any selected enzyme should leave a 4 bp overhang at the 5' end.

c) The selected enzymes in each set should all work efficiently under the same buffer conditions.

d) The selected enzymes in each set should ideally work efficiently at a single incubation temperature.

e) The chosen enzymes should be commercially available - ideally at concentrations of 10 U / µl or more.

f) The 5' overhangs left by any two enzymes in the same set should not be identical.

g) No enzyme should appear in both sets for TRSPA fabrication.

h) Enzymes should be selected to avoid or minimise the effects of mammalian methylation patterns. In particular, enzymes with CG dinucleotides in their recognition sites should be avoided unless the enzyme is known to be able to restrict ^{m5}CpG sites.

DNA methylation

In vertebrates, DNA is often methylated at the 5th position of cytosine in the sequence of CpG and this is the only chemical modification that DNA of vertebrates contains under physiological conditions. By the careful selection of enzymes which do not contain CpG sequences within the recognition site, or the selection of enzymes which freely restrict ^{m5}CpG methylated sites, it is possible to remove the potentially adverse effects of DNA methylation from the TRSPA analysis. 6 bp cutters which are known to restrict ^{m5}CpG modified DNA efficiently to leave a 5' four base overhang are *BspEI* and *XmaI*. These enzymes are therefore the only enzymes with restriction sites containing CpG dinucleotides that are potentially useful in a TRSPA analysis.

Enzyme selection method

Sixteen possible four base pair overhangs exist (excluding unusual enzymes with asymmetrical recognition sequences such as *BssSI* or *BsiI*), five of which contain CG in the sequence. A further four overhangs could potentially contain CG sequences within the restriction recognition site if preceded by a C and followed by a G. Enzymes are therefore preferentially selected from the remaining six groups.

Excluding isoschizomers, there are up to four possible enzymes which would leave a particular 5' four bp overhang. For example, enzymes leaving a CTAG overhang are

AvrII - CCTAGG

NheI - GCTAGC

SpeI - ACTAGT

XbaI - TCTAGA

Enzymes to cleave sites with all the combinations of flanking bases are not available for all overhangs - hence the enzyme choice is more limited for some overhang groups than others.

As a primary step towards enzyme selection, the enzymes supplied by Amersham Pharmacia Biotech, New England Biolabs and Promega are ordered below by overhang sequence. Supplementary details such as the percentage activity in common buffers, the reaction temperature, concentration and supplier are also recorded. For three overhang sequences, there are no available enzymes, the remaining 13 are described below. Enzymes considered unsuitable due to methylation sensitivity are shaded (darker shade). Enzymes considered unfavourable due to the presence of CpG sites even though they do restrict methylated DNA to some degree are also shaded (lighter shaded).

Candidate restriction enzymes leaving a 5' overhang of GTAC

Enzyme	Site	H/M/T buffer + BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
Acc65I	GGTACC	10/50/100	37	10	NEB	65 / 20
BsiWI	CGTACG	100(50)/100(50)/25(-)	55 (37)	12	NEB	80 / 20
BsrGI	TGTACA	10/100/100	37	10	NEB	80 / 20
SplI (BsiWI)	CGTACG	100/20/60	55	10	APB	N

Candidate restriction enzymes leaving a 5' overhang of TTAA

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
AflII	CTTAAG	25/100/100	37	10	APB	60 / 15

Candidate restriction enzymes leaving a 5' overhang of AATT

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
EcoRI	GAATTC	100/100/100	37	>40	APB/NEB	65 / 20
MfeI	CAATTG	10/50/100	37	10	NEB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of CATG

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
BspHI	TCATGA	50/100/100	37	10	NEB/APB	65 / 20
NcoI	CCATGG	100/100/100	37	50	NEB/APB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of GATC

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
BamHI	GGATCC	50/100/75	37	>40	APB/NEB	80 / 20
BclI	TGATCA	100/100/75	50 (37)	10	NEB	No
BglII	AGATCT	100/75/10	37	>40	APB/NEB	No

Candidate restriction enzymes leaving a 5' overhang of CCGG

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
BspEI	TCCGGA	100/10/0	37	10	NEB	80 / 20
NgoM IV	GCCGGC	10/50/100	37	10	NEB	80 / 20
XmaI	CCCGGG	0/100/100	37	10	NEB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of GCGC

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
KasI	GGCGCC	75/100/75	37	5	NEB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of TCGA

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
PaeR7I	CTCGAG	10/100/100	37	20	NEB	NB
SalI	GTCGAC	50/0/0	37	>40	APB/NEB	60 / 15
XhoI	CTCGAG	100/100/100	37	>40	APB/NEB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of AGCT

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
HindIII	AAGCTT	10/100/50	37	>40	NEB/APB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of CGCG

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
BssHII	GCGCGC	100/100/100	50	20	NEB/APB	80 / 20
MluI	ACGCGT	100/75/50	37	10	APB/NEB	65 / 20

Candidate restriction enzymes leaving a 5' overhang of GGCC

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
EagI	CGGCCG	100/25/10	37	50	NEB	65 / 20
Eco52I (EagI)	CGGCCG	20/20/20	37	10	APB	60 / 15

Candidate restriction enzymes leaving a 5' overhang of TGCA

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
ApaLI	GTGCAC	10/100/100	37	10	NEB/APB	70 / 15
Alw44I	GTGCAC	-/50/-	37	10	Promega	70 / 15

Candidate restriction enzymes leaving a 5' overhang of CTAG

Enzyme	Site	H/M/T buffer + BSA % activity	Optimu m °C	U / µl	Supplier	Inactivate °C/min
AvrII	CCTAGG	50/100/100	37	4	NEB	No
NheI	GCTAGC	10/100/100	37	10	APB/NEB	65 / 20
SpeI	ACTAGT	25/100/75	37	50	NEB	65 / 20
XbaI	TCTAGA	75/100/75	37	100	NEB	65 / 20
BlnI(AvrII)	CCTAGG	40/20/20	37	10	APB	No

From the above tables. a short-list of the most useful enzymes is given below for each of the buffer conditions shown

Overhang	Buffer H + BSA		Buffer M + BSA		Buffer T + BSA	
	Enzyme	Efficiency	Enzyme	Efficiency	Enzyme	Efficiency
AATT	*EcoRI*	100	*EcoRI*	100	*MfeI*	100
			MfeI	50	-	-
CATG	*NcoI*	100	*NcoI*	100	*NcoI*	100
	BspHI	50	*BspHI*	100	*BspHI*	100
GATC	*BglII*	100	*BamHI*	100	*BamHI*	75
	BamHI	50	*BglII*	75	-	-
TATA	-	-	-	-	-	-
ACGT	-	-	-	-	-	-
CCGG	BspEI	100	*XmaI*	100	*XmaI*	100
GCGC	-	-	-	-	-	-
TCGA	-	-	-	-	-	-
AGCT	-	-	*HindIII*	100	*HindIII*	50
CGCG	-	-	-	-	-	-
GGCC	-	-	-	-	-	-
TGCA	-	-	*ApaLI*	100	*ApaLI*	100
ATAT	-	-	-	-	-	-
CTAG	*XbaI*	75	*XbaI*	100	*NheI*	100
	AvrII	50	*SpeI*	100	*AvrII*	100
	BlnI	40	*NheI*	100	*XbaI*	75
	-	-	*AvrII*	100	*SpeI*	75
GTAC	*Acc65I*	100	*BsrGI*	100	*BsrGI*	100
			Acc65I	75	-	-
TTAA	-	-	*AflII*	100	*AflII*	100

Taking into account all of the criteria explained above, an

example selection of two six-enzyme sets is described below. Reserve enzymes, which could also be used, are shown as well. These reserve enzymes can be substituted (provided this does not cause overhang duplication) if practical problems regarding enzyme availability or performance should occur.

Example enzyme set 1 - for restriction in buffer M + BSA

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
1	*BamHI*	GGATCC	100	37	100	NEB	80 / 20
2	*BsrGI*	TGTACA	100	37	10	NEB	80 / 20
3	*HindIII*	AAGCTT	100	37	>40	APB	65 / 20
4	*NcoI*	CCATGG	100	37	50	NEB	65 / 20
5	*SpeI*	ACTAGT	100	37	50	NEB	65 / 20
6	*AflII*	CTTAAG	100	37	10	APB	60 / 15

Example enzyme set 2 - for restriction in buffer M + BSA

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
1	*EcoRI*	GAATTC	100	37	>40	APB	65 / 20
2	*BspHI*	TCATGA	100	37	10	APB	65 / 20
3	*BglII*	AGATCT	75	37	>40	APB	No
4	*XbaI*	TCTAGA	100	37	100	NEB	65 / 20
5	*Acc65I*	GGTACC	75	37	10	NEB	65 / 20
6	*ApaL1*	GTGCAC	100	37	10	NEB	No

1. Example reserve enzymes

Number	Enzyme	Site	M+BSA % activity	Optimum °C	U / µl	Supplier	Inactivate °C/min
r1	*XmaI*	CCCGGG	100	37	10	NEB	65 / 20
r2	*BspEI*	TCCGGA	100	37	10	NEB	80 / 20
r3	*NheI*	GCTAGC	100	37	10	APB	65/20
r4	*AvrII*	CCTAGG	100	37	4	NEB	No

Example specific (3x3) terminal restriction site profiling array (TRSPA) test matrices for above two sets of six enzymes

Triplet combinations from *BamHI*, *BsrGI*, *HindIII*, *NcoI*, *SpeI* and *AflII*

For example set 1 - *BamHI*, *BsrGI*, *HindIII*, *NcoI*, *SpeI* and *AflII* - the 20 (= ${}^{6}C_{3}$) triplet combinations are

BamHI BsrGI HindIII	*BamHI HindIII SpeI*	*BsrGI HindIII NcoI*	*BsrGI SpeI AflII*
BamHI BsrGI NcoI	*BamHI HindIII AflII*	*BsrGI HindIII SpeI*	*HindIII NcoI SpeI*
BamHI BsrGI SpeI	*BamHI NcoI SpeI*	*BsrGI HindIII AflII*	*HindIII NcoI AflII*
BamHI BsrGI AflII	*BamHI NcoI AflII*	*BsrGI NcoI SpeI*	*HindIII SpeI AflII*
BamHI HindIII NcoI	*BamHI SpeI AflII*	*BsrGI NcoI AflII*	*NcoI SpeI AflII*

Triplet combinations from *EcoRI*, *BspHI*, *BglII*, *XbaI*, *Acc65I* and *ApaLI*

For example set 2 - *EcoRI*, *BspHI*, *BglII*, *XbaI*, *Acc65I* and *ApaLI* - the 20 (= ${}^{6}C_{3}$) triplet combinations are

EcoRI BspHI BglII	*EcoRI BglII Acc65I*	*BspHI BglII XbaI*	*BspHI Acc65I ApaLI*
EcoRI BspHI XbaI	*EcoRI BglII ApaLI*	*BspHI BglII Acc65I*	*BglII XbaI Acc65I*
EcoRI BspHI Acc65I	*EcoRI XbaI Acc65I*	*BspHI BglII ApaLI*	*BglII XbaI ApaLI*
EcoRI BspHI ApaLI	*EcoRI XbaI ApaLI*	*BspHI XbaI Acc65I*	*BglII Acc65I ApaLI*
EcoRI BglII XbaI	*EcoRI Acc65I ApaLI*	*BspHI XbaI ApaLI*	*XbaI Acc65I ApaLI*

Example TRSPA-1 test matrices for set 1 - *BamHI*, *BsrGI*, *HindIII*, *NcoI*, *SpeI* and *AflII*

matrix 1	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*BsrGI*-primer	4	5	6
5'-biotin-*HindIII*-primer	7	8	9

matrix 2	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*NcoI*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*BsrGI*-primer	4	5	6
5'-biotin-*NcoI*-primer	7	8	9

matrix 3	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*SpeI*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*BsrGI*-primer	4	5	6
5'-biotin-*SpeI*-primer	7	8	9

matrix 4	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*AflII*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*BsrGI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 5	5'-HO-*BamHI*-primer	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*HindIII*-primer	4	5	6
5'-biotin-*NcoI*-primer	7	8	9

matrix 6	5'-HO-*BamHI*-primer	5'-HO-*HindIII*-primer	5'-HO-*SpeI*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*HindIII*-primer	4	5	6
5'-biotin-*SpeI*-primer	7	8	9

matrix 7	5'-HO-*BamHI*-primer	5'-HO-*HindIII*-primer	5'-HO-*AflII*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*HindIII*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 8	5'-HO-*BamHI*-primer	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*NcoI*-primer	4	5	6
5'-biotin-*SpeI*-primer	7	8	9

matrix 9	5'-HO-*BamHI*-primer	5'-HO-*NcoI*-primer	5'-HO-*AflII*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*NcoI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 10	5'-HO-*BamHI*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-biotin-*BamHI*-primer	1	2	3
5'-biotin-*SpeI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 11	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer
5'-biotin-*BsrGI*-primer	1	2	3
5'-biotin-*HindIII*-primer	4	5	6
5'-biotin-*NcoI*-primer	7	8	9

matrix 12	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer	5'-HO-*SpeI*-primer
5'-biotin-*BsrGI*-primer	1	2	3
5'-biotin-*HindIII*-primer	4	5	6
5'-biotin-*SpeI*-primer	7	8	9

matrix 13	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer	5'-HO-*AflII*-primer
5'-biotin-*BsrGI*-primer	1	2	3
5'-biotin-*HindIII*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 14	5'-HO-*BsrGI*-primer	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer
5'-biotin-*BsrGI*-primer	1	2	3
5'-biotin-*NcoI*-primer	4	5	6
5'-biotin-*SpeI*-primer	7	8	9

matrix 15	5'-HO-*BsrGI*-primer	5'-HO-*NcoI*-primer	5'-HO-*AflII*-primer
5'-biotin-*BsrGI*-primer	1	2	3
5'-biotin-*NcoI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 16	5'-HO-*BsrGI*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-biotin-*BsrGI*-primer	1	2	3
5'-biotin-*SpeI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 17	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer
5'-biotin-*HindIII*-primer	1	2	3
5'-biotin-*NcoI*-primer	4	5	6
5'-biotin-*SpeI*-primer	7	8	9

matrix 18	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer	5'-HO-*AflII*-primer
5'-biotin-*HindIII*-primer	1	2	3
5'-biotin-*NcoI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 19	5'-HO-*HindIII*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-biotin-*HindIII*-primer	1	2	3
5'-biotin-*SpeI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

matrix 20	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-biotin-*NcoI*-primer	1	2	3
5'-biotin-*SpeI*-primer	4	5	6
5'-biotin-*AflII*-primer	7	8	9

Example TRSPA-1 test matrices for set 2 - *EcoRI*, *BspHI*, *BglII*, *XbaI*, *Acc65I* and *ApaLI*

matrix 21	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BspHI*-primer	4	5	6
5'-biotin-*BglII*-primer	7	8	9

matrix 22	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*XbaI*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BspHI*-primer	4	5	6
5'-biotin-*XbaI*-primer	7	8	9

matrix 23	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*Acc65I*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BspHI*-primer	4	5	6
5'-biotin-*Acc65I*-primer	7	8	9

matrix 24	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BspHI*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 25	5'-HO-*EcoRI*-primer	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BglII*-primer	4	5	6
5'-biotin-*XbaI*-primer	7	8	9

matrix 26	5'-HO-*EcoRI*-primer	5'-HO-*BglII*-primer	5'-HO-*Acc65I*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BglII*-primer	4	5	6
5'-biotin-*Acc65I*-primer	7	8	9

matrix 27	5'-HO-*EcoRI*-primer	5'-HO-*BglII*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*BglII*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 28	5'-HO-*EcoRI*-primer	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*XbaI*-primer	4	5	6
5'-biotin-*Acc65I*-primer	7	8	9
matrix 29	5'-HO-*EcoRI*-primer	5'-HO-*XbaI*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*XbaI*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 30	5'-HO-*EcoRI*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*EcoRI*-primer	1	2	3
5'-biotin-*Acc65I*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 31	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer
5'-biotin-*BspHI*-primer	1	2	3
5'-biotin-*BglII*-primer	4	5	6
5'-biotin-*XbaI*-primer	7	8	9

matrix 32	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer	5'-HO-*Acc65I*-primer
5'-biotin-*BspHI*-primer	1	2	3
5'-biotin-*BglII*-primer	4	5	6
5'-biotin-*Acc65I*-primer	7	8	9

matrix 33	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*BspHI*-primer	1	2	3
5'-biotin-*BglII*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 34	5'-HO-*BspHI*-primer	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer
5'-biotin-*BspHI*-primer	1	2	3
5'-biotin-*XbaI*-primer	4	5	6
5'-biotin-*Acc65I*-primer	7	8	9

matrix 35	5'-HO-*BspHI*-primer	5'-HO-*XbaI*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*BspHI*-primer	1	2	3
5'-biotin-*XbaI*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 36	5'-HO-*BspHI*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*BspHI*-primer	1	2	3
5'-biotin-*Acc65I*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 37	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer
5'-biotin-*BglII*-primer	1	2	3
5'-biotin-*XbaI*-primer	4	5	6
5'-biotin-*Acc65I*-primer	7	8	9

matrix 38	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*BglII*-primer	1	2	3
5'-biotin-*XbaI*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 39	5'-HO-*BglII*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*BglII*-primer	1	2	3
5'-biotin-*Acc65I*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

matrix 40	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-biotin-*XbaI*-primer	1	2	3
5'-biotin-*Acc65I*-primer	4	5	6
5'-biotin-*ApaLI*-primer	7	8	9

Example TRSPA-2 test matrices for set 1 - *BamHI*, *BsrGI*, *HindIII*, *NcoI*, *SpeI* and *AflII*

matrix 1	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*BsrGI*-primer	4	5	6
5'-HO-*HindIII*-primer	7	8	9

matrix 2	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*NcoI*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*BsrGI*-primer	4	5	6
5'-HO-*NcoI*-primer	7	8	9

matrix 3	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*SpeI*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*BsrGI*-primer	4	5	6
5'-HO-*SpeI*-primer	7	8	9

matrix 4	5'-HO-*BamHI*-primer	5'-HO-*BsrGI*-primer	5'-HO-*AflII*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*BsrGI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 5	5'-HO-*BamHI*-primer	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*HindIII*-primer	4	5	6
5'-HO-*NcoI*-primer	7	8	9

matrix 6	5'-HO-*BamHI*-primer	5'-HO-*HindIII*-primer	5'-HO-*SpeI*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*HindIII*-primer	4	5	6
5'-HO-*SpeI*-primer	7	8	9

matrix 7	5'-HO-*BamHI*-primer	5'-HO-*HindIII*-primer	5'-HO-*AflII*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*HindIII*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 8	5'-HO-*BamHI*-primer	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*NcoI*-primer	4	5	6
5'-HO-*SpeI*-primer	7	8	9

matrix 9	5'-HO-*BamHI*-primer	5'-HO-*NcoI*-primer	5'-HO-*AflII*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*NcoI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 10	5'-HO-*BamHI*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-HO-*BamHI*-primer	1	2	3
5'-HO-*SpeI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 11	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer
5'-HO-*BsrGI*-primer	1	2	3
5'-HO-*HindIII*-primer	4	5	6
5'-HO-*NcoI*-primer	7	8	9

matrix 12	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer	5'-HO-*SpeI*-primer
5'-HO-*BsrGI*-primer	1	2	3
5'-HO-*HindIII*-primer	4	5	6
5'-HO-*SpeI*-primer	7	8	9

matrix 13	5'-HO-*BsrGI*-primer	5'-HO-*HindIII*-primer	5'-HO-*AflII*-primer
5'-HO-*BsrGI*-primer	1	2	3
5'-HO-*HindIII*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 14	5'-HO-*BsrGI*-primer	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer
5'-HO-*BsrGI*-primer	1	2	3
5'-HO-*NcoI*-primer	4	5	6
5'-HO-*SpeI*-primer	7	8	9

matrix 15	5'-HO-*BsrGI*-primer	5'-HO-*NcoI*-primer	5'-HO-*AflII*-primer
5'-HO-*BsrGI*-primer	1	2	3
5'-HO-*NcoI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 16	5'-HO-*BsrGI*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-HO-*BsrGI*-primer	1	2	3
5'-HO-*SpeI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 17	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer
5'-HO-*HindIII*-primer	1	2	3
5'-HO-*NcoI*-primer	4	5	6
5'-HO-*SpeI*-primer	7	8	9

matrix 18	5'-HO-*HindIII*-primer	5'-HO-*NcoI*-primer	5'-HO-*AflII*-primer
5'-HO-*HindIII*-primer	1	2	3
5'-HO-*NcoI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 19	5'-HO-*HindIII*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-HO-*HindIII*-primer	1	2	3
5'-HO-*SpeI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

matrix 20	5'-HO-*NcoI*-primer	5'-HO-*SpeI*-primer	5'-HO-*AflII*-primer
5'-HO-*NcoI*-primer	1	2	3
5'-HO-*SpeI*-primer	4	5	6
5'-HO-*AflII*-primer	7	8	9

Example TRSPA-2 test matrices for set 2 - *EcoRI*, *BspHI*, *BglII*, *XbaI*, *Acc65I* and *ApaLI*

matrix 21	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BspHI*-primer	4	5	6
5'-HO-*BglII*-primer	7	8	9

matrix 22	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*XbaI*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BspHI*-primer	4	5	6
5'-HO-*XbaI*-primer	7	8	9

matrix 23	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*Acc65I*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BspHI*-primer	4	5	6
5'-HO-*Acc65I*-primer	7	8	9

matrix 24	5'-HO-*EcoRI*-primer	5'-HO-*BspHI*-primer	5'-HO-*ApaLI*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BspHI*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 25	5'-HO-*EcoRI*-primer	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BglII*-primer	4	5	6
5'-HO-*XbaI*-primer	7	8	9

matrix 26	5'-HO-*EcoRI*-primer	5'-HO-*BglII*-primer	5'-HO-*Acc65I*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BglII*-primer	4	5	6
5'-HO-*Acc65I*-primer	7	8	9

matrix 27	5'-HO-*EcoRI*-primer	5'-HO-*BglII*-primer	5'-HO-*ApaLI*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*BglII*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 28	5'-HO-*EcoRI*-primer	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*XbaI*-primer	4	5	6
5'-HO-*Acc65I*-primer	7	8	9

matrix 29	5'-HO-*EcoRI*-primer	5'-HO-*XbaI*-primer	5'-HO-*ApaLI*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*XbaI*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 30	5'-HO-*EcoRI*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-HO-*EcoRI*-primer	1	2	3
5'-HO-*Acc65I*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 31	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer
5'-HO-*BspHI*-primer	1	2	3
5'-HO-*BglII*-primer	4	5	6
5'-HO-*XbaI*-primer	7	8	9

matrix 32	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer	5'-HO-*Acc65I*-primer
5'-HO-*BspHI*-primer	1	2	3
5'-HO-*BglII*-primer	4	5	6
5'-HO-*Acc65I*-primer	7	8	9

matrix 33	5'-HO-*BspHI*-primer	5'-HO-*BglII*-primer	5'-HO-*ApaLI*-primer
5'-HO-*BspHI*-primer	1	2	3
5'-HO-*BglII*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 34	5'-HO-*BspHI*-primer	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer
5'-HO-*BspHI*-primer	1	2	3
5'-HO-*XbaI*-primer	4	5	6
5'-HO-*Acc65I*-primer	7	8	9

matrix 35	5'-HO-*BspHI*-primer	5'-HO-*XbaI*-primer	5'-HO-*ApaLI*-primer
5'-HO-*BspHI*-primer	1	2	3
5'-HO-*XbaI*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 36	5'-HO-*BspHI*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-HO-*BspHI*-primer	1	2	3
5'-HO-*Acc65I*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 37	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer
5'-HO-*BglII*-primer	1	2	3
5'-HO-*XbaI*-primer	4	5	6
5'-HO-*Acc65I*-primer	7	8	9

matrix 38	5'-HO-*BglII*-primer	5'-HO-*XbaI*-primer	5'-HO-*ApaLI*-primer
5'-HO-*BglII*-primer	1	2	3
5'-HO-*XbaI*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 39	5'-HO-*BglII*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-HO-*BglII*-primer	1	2	3
5'-HO-*Acc65I*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

matrix 40	5'-HO-*XbaI*-primer	5'-HO-*Acc65I*-primer	5'-HO-*ApaLI*-primer
5'-HO-*XbaI*-primer	1	2	3
5'-HO-*Acc65I*-primer	4	5	6
5'-HO-*ApaLI*-primer	7	8	9

Hybridisation patterns

There are six possible hybridisation patterns for a given probe fragment from inter-population perfectly matched duplex depletion enrichment with a (3x3) test matrix in TRSPA-1 analysis.

There are also six possible hybridisation patterns for a given probe fragment from inter-population perfectly matched duplex depletion enrichment with a (3x3) test matrix in TRSPA-2 analysis.

We can denote these patterns as follows

TRSPA-1 analysis

Pattern a (1 only)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-biotin-X-primer	1	2	3
5'-biotin-Y-primer	4	5	6
5'-biotin-Z-primer	7	8	9

Pattern b (2 and 4)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-biotin-X-primer	1	2	3
5'-biotin-Y-primer	4	5	6
5'-biotin-Z-primer	7	8	9

Pattern c (3 and 7)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-biotin-X-primer	1	2	3
5'-biotin-Y-primer	4	5	6
5'-biotin-Z-primer	7	8	9

Pattern d (5 only)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-biotin-X-primer	1	2	3
5'-biotin-Y-primer	4	5	6
5'-biotin-Z-primer	7	8	9

Pattern e (6 and 8)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-biotin-X-primer	1	2	3
5'-biotin-Y-primer	4	5	
5'-biotin-Z-primer	7		9

Pattern f (9 only)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-biotin-X-primer	1	2	3
5'-biotin-Y-primer	4	5	6
5'-biotin-Z-primer	7	8	

TRSPA-2 analysis

Pattern a (row 1 and column 1)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'-HO-X-primer			
5'-HO-Y-primer		5	6
5'-HO-Z-primer		8	9

Pattern b (2 and 4)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'- HO -X-primer	1		3
5'- HO -Y-primer		5	6
5'- HO -Z-primer	7	8	9

Pattern c (3 and 7)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'- HO -X-primer	1	2	
5'- HO -Y-primer	4	5	6
5'- HO -Z-primer		8	9

Pattern d (row 2 and column 2)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'- HO -X-primer	1		3
5'- HO -Y-primer			
5'- HO -Z-primer	7		9

Pattern e (6 and 8)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'- HO -X-primer	1	2	3
5'- HO -Y-primer	4	5	
5'- HO -Z-primer	7		9

Pattern f (row 3 and column 3)

matrix #	5'-HO-X-primer	5'-HO-Y-primer	5'-HO-Z-primer
5'- HO -X-primer	1	2	
5'- HO -Y-primer	4	5	
5'- HO -Z-primer			

Signatures

If we have 40 such (3x3) test matrices and we denote the possible hybridisation patterns a, b, c, d, e and f, then we can write the overall TRSPA hybridisation signature as follows

Matrix	Pattern	matrix	pattern
1	a. b, c. d. e or f	21	a, b, c, d. e or f
2	a. b, c, d. e or f	22	a, b, c. d, e or f
3	a, b. c, d, e or f	23	a, b, c, d, e or f
4	a, b, c. d. e or f	24	a, b, c, d, e or f
5	a, b, c. d, e or f	25	a, b, c, d, e or f
6	a, b, c, d, e or f	26	a, b, c, d, e or f
7	a. b, c, d. e or f	27	a, b, c, d, e or f
8	a, b, c, d, e or f	28	a, b, c, d, e or f
9	a, b, c, d, e or f	29	a, b, c, d. e or f
10	a. b, c, d, e or f	30	a. b, c, d. e or f
11	a, b, c. d. e or f	31	a, b, c, d, e or f
12	a, b, c, d, e or f	32	a, b, c, d, e or f
13	a, b, c, d, e or f	33	a, b, c, d, e or f
14	a, b, c, d, e or f	34	a, b, c, d, e or f
15	a, b, c. d, e or f	35	a, b, c, d, e or f
16	a, b, c, d, e or f	36	a, b, c, d, e or f
17	a, b, c, d, e or f	37	a, b, c, d, e or f
18	a, b, c, d, e or f	38	a, b, c, d, e or f
19	a, b, c, d. e or f	39	a, b, c, d, e or f
20	a. b. c, d, e or f	40	a. b, c, d, e or f

Aspect (III)

In another aspect the invention provides a nucleic acid characterisation method which comprises presenting to the set of arrays as defined above a nucleic acid fragment of interest under hybridisation conditions, and observing a pattern of hybridisation. Preferably, a plurality of nucleic acid fragments of interest are separately presented to the set of arrays, and the resulting patterns of hybridisation are compared. Preferably, the plurality of nucleic acid fragments of interest are drawn from the mixture of DNA fragments, enriched in fragments that are characteristic of a phenotype of interest, as described under the invention (1) above.

Thus in this aspect the invention provides a method of identifying fragments of DNA that are characteristic of a phenotype of

interest, which method comprises recovering, cloning and amplifying individual DNA fragments from the mixture of DNA fragments obtained under invention (1) above, presenting the individual DNA fragments to the set of arrays as defined under hybridisation conditions, observing a pattern of hybridisation generated by each individual DNA fragment, and subjecting to further investigation any two or more individual DNA fragments whose hybridisation patterns are similar or identical.

TRSPA signatures and whole genome association studies

After a given number of cycles of inter-population perfectly matched duplex depletion, phenotype determining fragments will be enriched but will not be entirely free from 'noise' fragments. Noise may result from unequal allelic frequencies for certain SNPs between the two populations. Noise will also result from the presence of somatic mutations in the cells used to prepare DNA fragments and from the use of polymerase chain reaction in some of the embodiments of the current invention.

For the preferred embodiment, DNA is prepared from a library of clones (either genomic clones or cDNA clones) - with inserts derived from the individual(s) and propagated in some appropriate host and cloning vector system. Restriction enzyme fragmentation is used prior to cloning and polymerase chain reaction amplification is used to prepare the DNA for comparison in fragmented form. Priming sites within the vector sequence flanking the cloned restriction enzyme fragmented inserts are employed for one or more cycles of polymerase chain reaction amplification of the fragmented DNA of interest. The primers used for polymerase chain reaction amplification of the fragmented DNA of interest are again used after the phenotype-determining fragment enrichment process to 'rescue' and clone the enriched fragments. Cloned enriched fragments are colony purified, picked into appropriate storage containers, catalogued and archived, DNA probes are prepared from these single clones and are

individually hybridised against 180 spot TRSPA arrays as above. TRSPA signatures are determined for many colonies. Most noise fragments will be random in nature and will thus be randomly distributed amongst the 6^{20} = $3.7 \cdot 10^{15}$ possible types of TRSPA signature. Phenotype-determining signal fragments, however, will be those where repeat TRSPA signatures are obtained as more and more colonies are sampled. The more frequently the repetition of a given TRSPA signature occurs per unit colonies sampled, the greater the signal to noise ratio and the more successful has been the enrichment.

A great many of the steps in this process are amenable to high throughput automation - enabling very large numbers of single colony TRSPA signatures to be determined with ease and extending the power of the current invention to cases where signal to noise ratios are beyond current approaches.

Statistical correlations (associations) can initially be drawn between TRSPA signatures and phenotype. The clones giving rise to a particular TRSPA signature showing a useful association with a phenotype of interest can then be sequenced in order to determine at a DNA sequence level the association(s) with the phenotype of interest. Such associations have future predictive values for the phenotype of interest, knowing the genotype and will be of great value in medicine and pharmacogenetics.

If the genome of interest is wholly or partially sequenced, we can also *in silico* restrict the DNA with all n enzymes, calculate the expected signature for each fragment and pattern match these expected signatures with the observed signature (taking into account any loss or gain of restriction sites due to polymorphic variation compared to the reference sequence) to immediately identify the fragment of interest within a gene, genomic region, chromosome or whole genome. This latter method will be of great value in those cases where a great many phenotype-determining fragments are obtained and repeat signatures are rare or unobtained. The

clustering of phenotype-determining fragments to adjacent DNA regions thus gives an association between those genomic regions and the phenotype of interest.

Aspect (IV)

In yet another aspect the invention provides a double-stranded DNA molecule having the sequence a-A-b-B...X-y-Y-z where A, B...X and Y are unique restriction sites for n different restriction endonuclease enzymes, and a, b...y, z denotes distances in base pairs, characterised in that each fragment, obtainable by cutting the DNA molecule by means of any one or more up to n of the restriction enzymes, has a different length from every other fragment.

An example totally diagnostic internal control DNA which allows both the extent and exact nature of any example set 1 (or example set 2) 6 bp cutter partial digestion to be unambiguously determined for inter-population perfectly matched duplex depletion or TRSPA restriction

In both of the above schemes, it is important that limit digestion products are obtained. Monitoring the extent of partial digestion resulting from multi-enzymatic restriction and determining precisely which enzymes have failed to cut is a task of great importance.

If we have up to six enzymes for DNA digestion - let us label these A, B, C, D, E and F. We need to somehow determine that these have all cut to completion during the fragmentation stage for inter-population perfectly matched duplex depletion and also for the digestion step prior to adaptor ligation in TRSPA fabrication. If any of the enzymes have failed to cut to completion, we need to know which ones and to what degree in order to effectively rectify the problem.

The structure of th internal control DNA

If we construct a double stranded DNA molecule with the

following structure

end---t---A---u---B---v---C---w---D---x---E---y---F---z---end

where the A, B, C, D, E and F denote the sites for restriction enzyme cutting and t, u, v, w, x, y and z denote distances in base pairs.

This internal control DNA is either uniformly pre-labelled and added to the DNA of interest at an appropriate concentration prior to restriction or is Southern blot probed with a complementary sequence not found in the DNA of interest after restriction.

All six enzymes can cut in only one way.

One enzyme can fail to cut in ${}^{6}C_{1} = 6$ ways, these are: A, B, C, D, E or F failing to cut.

Two enzymes can fail to cut in ${}^{6}C_{2} = 15$ ways, these are: AB, AC, AD, AE, AF, BC, BD, BE, BF, CD, CE, CF, DE, DF or EF failing to cut.

Three enzymes can fail to cut in ${}^{6}C_{3} = 20$ ways, these are: ABC, ABD, ABE, ABF, ACD, ACE, ACF, ADE, ADF, AEF, BCD, BCE, BCF, BDE, BDF, BEF, CDE, CDF, CEF or DEF failing to cut.

Four enzymes can fail to cut in ${}^{6}C_{4} = 15$ ways, these are: ABCD, ABCE, ABCF, ABDE, ABDF, ABEF, ACDE, ACDF, ACEF, ADEF, BCDE, BCDF, BCEF, BDEF or CDEF failing to cut.

Five enzymes can fail to cut in ${}^{6}C_{5} = 6$ ways, these are: ABCDE, ABCDF, ABCEF, ABDEF, ACDEF or BCDEF failing to cut.

All six enzymes can fail to cut in only one way.

Each of the above possibilities will generate one or more fragments from the internal control DNA. If each possible fragment has a discernible size from any other, then we can determine exactly which enzymes have cut and which have not from the size distribution of the fragments generated. The task is therefore to design such a DNA molecule.

Example simulations

Seven simulations are given below - varying the size of inter-site fragments. Criteria for a successful outcome include the following

1. The inter-fragment spacing should be greater for larger fragments (so as to aid electrophoretic resolution).

2. All possible fragments should be unambiguously resolvable in size from each-other.

3. Size gaps between bands comprising different numbers of inter-site units should be greater than the size gaps between bands comprising the same number of inter-site units.

4. The size gaps and size spread from largest to smallest fragment should be electrophoretically compatible.

Simulation 1

Inter-site fragment sizes (in bp)

End---A	A---B	B---C	C---D	D---E	E---F	F---end
80	90	100	110	120	130	140

Possible digestion products obtained (in bp)

	Fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	80	
A---B	90	10
B---C	100	10
C---D	110	10
D---E	120	10
E---F	130	10
F---end	140	10
two unit fragments		
end---B	170	30
A---C	190	20
B---D	210	20
C---E	230	20
D---F	250	20
E---end	270	20
three unit fragments		
end---C	270	0
A---D	300	30
B---E	330	30
C---F	360	30
D---end	390	30
four unit fragments		
end---D	380	-10
A---E	420	40
B---F	460	40
C---end	500	40
five unit fragments		
end---E	500	0
A---F	550	50
B---end	600	50
six unit fragments		
end---F	630	30
B---end	690	60
seven unit fragments		
end---end	770	80

Spread = 690 bp

Simulation 2

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
90	100	110	120	130	140	150

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	90	
A---B	100	10
B---C	110	10
C---D	120	10
D---E	130	10
E---F	140	10
F---end	150	10
two unit fragments		
end---B	190	40
A---C	210	20
B---D	230	20
C---E	250	20
D---F	270	20
E---end	290	20
three unit fragments		
end---C	300	10
A---D	330	30
B---E	360	30
C---F	390	30
D---end	420	30
four unit fragments		
end---D	420	0
A---E	460	40
B---F	500	40
C---end	540	40
five unit fragments		
end---E	550	10
A---F	600	50
B---end	650	50
six unit fragments		
end---F	690	40
B---end	750	60
seven unit fragments		
end---end	840	90

Spread = 750 bp

Simulation 3

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
100	110	120	130	140	150	160

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	100	
A---B	110	10
B---C	120	10
C---D	130	10
D---E	140	10
E---F	150	10
F---end	160	10
two unit fragments		
end---B	210	50
A---C	230	20
B---D	250	20
C---E	270	20
D---F	290	20
E---end	310	20
three unit fragments		
end---C	330	20
A---D	360	30
B---E	390	30
C---F	420	30
D---end	450	30
four unit fragments		
end---D	460	10
A---E	500	40
B---F	540	40
C---end	580	40
five unit fragments		
end---E	600	20
A---F	650	50
B---end	700	50
six unit fragments		
end---F	750	50
B---end	810	60
seven unit fragments		
end---end	910	100

Spread = 810 bp

Simulation 4

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
110	120	130	140	150	160	170

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	110	
A---B	120	10
B---C	130	10
C---D	140	10
D---E	150	10
E---F	160	10
F---end	170	10
two unit fragments		
end---B	230	60
A---C	250	20
B---D	270	20
C---E	290	20
D---F	310	20
E---end	330	20
three unit fragments		
end---C	360	30
A---D	390	30
B---E	420	30
C---F	450	30
D---end	480	30
four unit fragments		
end---D	500	20
A---E	540	40
B---F	580	40
C---end	620	40
five unit fragments		
end---E	650	30
A---F	700	50
B---end	750	50
six unit fragments		
end---F	810	60
B---end	870	60
seven unit fragments		
end---end	980	110

Spread = 870 bp

Simulation 5

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
120	130	140	150	160	170	180

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	120	
A---B	130	10
B---C	140	10
C---D	150	10
D---E	160	10
E---F	170	10
F---end	180	10
two unit fragments		
end---B	250	70
A---C	270	20
B---D	290	20
C---E	310	20
D---F	330	20
E---end	350	20
three unit fragments		
end---C	390	40
A---D	420	30
B---E	450	30
C---F	480	30
D---end	510	30
four unit fragments		
end---D	540	30
A---E	580	40
B---F	620	40
C---end	660	40
five unit fragments		
end---E	700	40
A---F	750	50
B---end	800	50
six unit fragments		
end---F	870	70
B---end	930	60
seven unit fragments		
end---end	1050	120

Spread = 930 bp

Simulation 6

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
130	140	150	160	170	180	190

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	130	
A---B	140	10
B---C	150	10
C---D	160	10
D---E	170	10
E---F	180	10
F---end	190	10
two unit fragments		
end---B	270	80
A---C	290	20
B---D	310	20
C---E	330	20
D---F	350	20
E---end	370	20
three unit fragments		
end---C	420	50
A---D	450	30
B---E	480	30
C---F	510	30
D---end	540	30
four unit fragments		
end---D	580	40
A---E	620	40
B---F	660	40
C---end	700	40
five unit fragments		
end---E	750	50
A---F	800	50
B---end	850	50
six unit fragments		
end---F	930	80
B---end	990	60
seven unit fragments		
end---end	1120	130

Spread = 990 bp

Simulation 7

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
140	150	160	170	180	190	200

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	140	
A---B	150	10
B---C	160	10
C---D	170	10
D---E	180	10
E---F	190	10
F---end	200	10
two unit fragments		
end---B	290	90
A---C	310	20
B---D	330	20
C---E	350	20
D---F	370	20
E---end	390	20
three unit fragments		
end---C	450	60
A---D	480	30
B---E	510	30
C---F	540	30
D---end	570	30
four unit fragments		
end---D	620	50
A---E	660	40
B---F	700	40
C---end	740	40
five unit fragments		
end---E	800	60
A---F	850	50
B---end	900	50
six unit fragments		
end---F	990	90
B---end	1050	60
seven unit fragments		
end---end	1190	140

Spread = 1050 bp

According to the above criteria for success, simulation 7 above clearly fulfils all of the requirements.

An example totally diagnostic internal control DNA which allows both the extent and exact nature of any example set 1 (or example set 2) 4 bp cutter partial digestion to be unambiguously determined for inter-population perfectly matched duplex depletion

If we have up to three enzymes for DNA digestion - let us label these A, B and C. We need to somehow determine that these have all cut to completion during the fragmentation stage for inter-population perfectly matched duplex depletion. If any of the enzymes have failed to cut to completion, we need to know which ones and to what degree in order to effectively rectify the problem.

The structure of the internal control DNA

If we construct a double stranded DNA molecule with the following structure

end---t---A---u---B---v---C---w---end

where the A, B and C denote the sites for restriction enzyme cutting and t, u, v and w denote distances in base pairs.

This internal control DNA is uniformly pre-labelled and added to the DNA of interest at an appropriate concentration prior to restriction or is Southern blot probed with a complementary sequence not found in the DNA of interest after restriction.

All three enzymes can cut in only one way.

One enzyme can fail to cut in ${}^3C_1 = 3$ ways, these are: A, B or C failing to cut.

Two enzymes can fail to cut in ${}^3C_2 = 3$ ways, these are: AB, AC or BC failing to cut.

All three enzymes can fail to cut in only one way.

Each of the above possibilities will generate one or more fragments from the internal control DNA. If each possible fragment has a discernible size from any other (and from any of the fragments in simulation 7 above for the panel of up to 6 enzymes), then we can determine exactly which enzymes have cut and which have not from the size distribution of the fragments generated. The task is therefore to design such a DNA molecule.

Example simulations

Six simulations are given below - varying the size of inter-site fragments. Criteria for a successful outcome included the following:

1. The inter-fragment spacing should be greater for larger fragments (so as to aid electrophoretic resolution).
2. All possible fragments should be unambiguously resolvable in size from each-other.
3. Size gaps between bands comprising different numbers of inter-site units should ideally be greater than the size gaps between bands comprising the same number of inter-site units.
4. The size gaps and size spread from largest to smallest fragment should be electrophoretically compatible.
5. The largest fragment obtained should ideally be smaller than the smallest fragment obtained in simulation 7 above for the panel of up to six enzymes.

Simulation 1

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
10	20	30	40

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	10	
A---B	20	10
B---C	30	10
C---end	40	10
two unit fragments		
end---B	30	-10
A---C	50	20
B---end	70	20
three unit fragments		
end---C	60	-10
A---end	90	30
four unit fragments		
end---end	100	10

Spread = 90 bp

Simulation 2

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
15	25	35	45

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	15	
A---B	25	10
B---C	35	10
C---end	45	10
two unit fragments		
end---B	40	-5
A---C	60	20
B---end	80	20
three unit fragments		
end---C	75	-5
A---end	105	30
four unit fragments		
end---end	120	15

Spread = 105 bp

Simulation 3

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
20	30	40	50

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	20	
A---B	30	10
B---C	40	10
C---end	50	10
two unit fragments		
end---B	50	0
A---C	70	20
B---end	90	20
three unit fragments		
end---C	90	0
A---end	120	30
four unit fragments		
end---end	140	20

Spread = 120 bp

Simulation 4

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
25	35	45	55

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	25	
A---B	35	10
B---C	45	10
C---end	55	10
two unit fragments		
end---B	60	5
A---C	80	20
B---end	100	20
three unit fragments		
end---C	105	5
A---end	135	30
four unit fragments		
end---end	160	25

Spread = 135 bp

Simulation 5

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
20	25	30	35

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	20	
A---B	25	5
B---C	30	5
C---end	35	5
two unit fragments		
end---B	45	10
A---C	55	10
B---end	65	10
three unit fragments		
end---C	75	10
A---end	90	15
four unit fragments		
end---end	110	20

Spread = 90 bp

Simulation 6

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
25	30	35	40

Possible digestion products obtained (in bp)

	fragment length (in bp)	size gap to the next smallest fragment (in bp)
one unit fragments		
end---A	25	
A---B	30	5
B---C	35	5
C---end	40	5
two unit fragments		
end---B	55	15
A---C	65	10
B---end	75	10
three unit fragments		
end---C	90	15
A---end	105	15
four unit fragments		
end---end	130	25

Spread = 105 bp

According to the above criteria for success, simulation 6 clearly fulfils all of the requirements.

Example determination of the entire set of internal control DNA limit and partial digestion patterns for a panel of up to six restriction enzymes

For the example simulation 7, the entire set of internal control DNA limit and partial digestion patterns for a panel of up to six restriction enzymes can be determined as below.

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---D	D---E	E---F	F---end
140	150	160	170	180	190	200

Possible digestion products obtained

All six enzymes can cut in only one way.

failed digests	bp	bp	bp	bp	bp	bp	bp
none	140	150	160	170	180	190	200

One enzyme can fail to cut in ${}^{6}C_{1} = 6$ ways, these are: A, B, C, D, E or F failing to cut.

failed digest	bp	bp	bp	bp	bp	bp
A	290	160	170	180	190	200
B	140	310	170	180	190	200
C	140	150	330	180	190	200
D	140	150	160	350	190	200
E	140	150	160	170	370	200
F	140	150	160	170	180	390

Two enzymes can fail to cut in ${}^{6}C_{2}$ = 15 ways, these are: AB, AC, AD, AE, AF, BC, BD, BE, BF, CD, CE, CF, DE, DF or EF failing to cut.

failed digests	bp	bp	bp	bp	bp
AB	450	170	180	190	200
AC	290	330	180	190	200
AD	290	160	350	190	200
AE	290	160	170	370	200
AF	290	160	170	180	390
BC	140	480	180	190	200
BD	140	310	350	190	200
BE	140	310	170	370	200
BF	140	310	170	180	390
CD	140	150	510	190	200
CE	140	150	330	370	200
CF	140	150	330	180	390
DE	140	150	160	540	200
DF	140	150	160	350	390
EF	140	150	160	170	570

5

Three enzymes can fail to cut in ${}^{6}C_{3}$ = 20 ways, these are: ABC, ABD, ABE, ABF, ACD, ACE, ACF, ADE, ADF, AEF, BCD, BCE, BCF, BDE, BDF, BEF, CDE, CDF, CEF or DEF failing to cut.

failed digests	bp	bp	bp	bp
ABC	620	180	190	200
ABD	450	350	190	200
ABE	450	170	370	200
ABF	450	170	180	390
ACD	290	510	190	200
ACE	290	330	370	200
ACF	290	330	180	390
ADE	290	160	540	200
ADF	290	160	350	390
AEF	290	160	170	570
BCD	140	660	190	200
BCE	140	480	370	200
BCF	140	480	180	390
BDE	140	310	540	200
BDF	140	310	350	390
BEF	140	310	170	570
CDE	140	150	700	200
CDF	140	150	510	390
CEF	140	150	330	570
DEF	140	150	160	740

Four enzymes can fail to cut in ${}^{6}C_{4}$ = 15 ways, these are: ABCD, ABCE, ABCF, ABDE, ABDF, ABEF, ACDE, ACDF, ACEF, ADEF, BCDE, BCDF, BCEF, BDEF or CDEF failing to cut.

failed digests	bp	bp	bp
ACCD	800	190	200
ABCE	620	370	200
ABCF	620	180	390
ABDE	450	540	200
ABDF	450	350	390
ABEF	450	170	570
ACDE	290	700	200
ACDF	290	510	390
ACEF	290	330	570
ADEF	290	160	740
BCDE	140	850	200
BCDF	140	660	390
BCEF	140	480	570
BDEF	140	310	740
CDEF	140	150	900

Five enzymes can fail to cut in ${}^{6}C_{5} = 6$ ways, these are: ABCDE, ABCDF, ABCEF, ABDEF, ACDEF or BCDEF failing to cut.

failed digests	bp	bp
ABCDE	990	200
ABCDF	800	390
ABCEF	620	570
ABDEF	450	740
ACDEF	290	900
BCDEF	140	1050

All six enzymes can fail to cut in only one way.

failed digests	bp
all	1190

Example determination of the entire set of internal control DNA limit and partial digestion patterns for a panel of up to three restriction enzymes

For the example simulation 6, the entire set of internal control DNA limit and partial digestion patterns for a panel of up to three restriction enzymes can be determined as below.

Inter-site fragment sizes (in bp)

end---A	A---B	B---C	C---end
25	30	35	40

Possible digestion products obtained

All three enzymes can cut in only one way.

failed digests	bp	bp	bp	bp
none	25	30	35	40

One enzyme can fail to cut in $^3C_1 = 3$ ways. these are: A, B or C failing to cut.

failed digests	bp	bp	bp
A	55	35	40
B	25	65	40
C	25	30	75

Two enzymes can fail to cut in $^3C_2 = 3$ ways, these are: AB, AC or BC failing to cut.

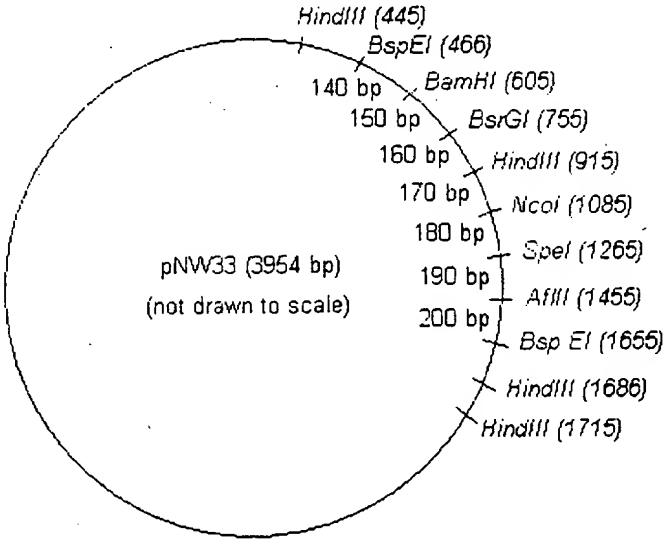
failed digests	bp	bp
AB	90	40
AC	55	75
BC	25	105

All three enzymes can fail to cut in only one way.

failed digests	bp
all	130

Example 1a – The digestion internal control plasmid for the 6 bp cutter set 1 TRSPA enzymes *BamHI, BsrGI, HindIII, NcoI, SpeI,* and *AflII*

The plasmid pNW33 (shown below) was constructed to contain an insert with all of the 6 bp cutter TRSPA enzyme sites.



BspEI sites define the outer ends of the 140 bp and the 200 bp fragments. The full sequence for pNW33 is shown below:

tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgcc
gggagcagacaagcccgtcagggcgcgtcagcgggtgttggcgggtgtcggggctggcttaactatgcggcatcagagca
gattgtactgagagtgcaccatatgcggtgtgaaataccgcacagatgcgtaaggagaaaataccgcatcaggcgccattc
gccattcaggctgcgcaactgttgggaagggcgatcggtgcgggcctcttcgctattacgccagctggcgaaaggggatgt
gctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgtaaaacgacggccagtgaattcgagctcgg
taccgggccccccctcgaggtcgacggtatcgataagcttgatcgcagctggtaatccggacgcccgcgtcgaagatgttgg
ggtgttgtaacaatatcgattccaattcagcgggggccacctgatatcctttgtatttaattaaagacttcaagcggtcaactatga
agaagtgttcgtcttcgtcccagtaaggatccgcactttgaattttgtaatcctgaagggatcgtaaaaacagctcttcttcaaatct
atacattaagacgactcgaaatccacatatcaaatatccgagtgtagtaaacattccaaaaccgtgatggaatggaacaaca
cttaaatgtacaccctggtaatccgttttagaatccatgataataattttctggattattggtaatttttttgcacgttcaaaatttttgca
acccctttttggaaacaaacactacggtaggctgcgaaatgttcatactgttgagcaattcacgttcattataagcttttcactgcat
acgacgattctgtgatttgtattcagcccatatcgtttcatagcttctgccaaccgaacggacatttcgaagtattccgcgtacgtg

atgttcacctcgatatgtgcatctgtaaaagcaattgttccaggaaccagggcgtatctcttcatagccatggaatacgccttttc
agtgttgcgatgctaatgccgttacaaatattccgagcaccaagaatggctgcgcgcttgcctggtacttgacgtcgtatttgacg
gggtccttgagaaagtatttaaactggaacacaatctgaggaatgatcaaagcaaccaacgccaacgcataataactagtg
caataccaagacctcccaataatagcacccagacttgtgtaataacctctggctctgatattgctccagatggaattggacgat
atggctcattaattgcgtcgatatctctatcataccagtcgttgattgtctgtgtatagccagtaagacaaggaccagacatcatca
tgcaaagaatcgcttaagcccttcttggcctttatgaggatctctctgatttttcttgcgtcgagtttccggtaagacctttcggtactt
cgtccacaaacacaactcctccgcgcaacttttcgcggttgttacttgactggccacgtaatccacgatctcttttccgtcatcgt
ctttccgtgctccaaaacaacaacggcggcgggtccggattaccagctgcgatcaagcttatcgataccgtcgacctcgacct
gcaggcatgcaagcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacga
gccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttc
cagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttc
cgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggtt
atccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaag
gccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaa
acccgacaggactataaagataccaggcgtttcccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccg
gatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcg
ctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaaccc
ggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacag
agttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcgga
aaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgc
agaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggatt
ttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagt
aaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgact
ccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctca
ccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctcc
atccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggc
atcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgt
gcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagc
actgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgta
tgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcatt
ggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaac
tgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataag
ggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggataca
tatttgaatgtatttagaaaaataaacaaatagggggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaaacc
attattatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtc

The insert was introduced as a four fragment ligation of 140 (KpnI-BamHI), 150 / 160 (BamHI-HindIII), 170 / 180 (HindIII-SpeI) and 190 / 200 (SpeI-XhoI) into KpnI / SalI digested pUC19c DNA (Genbank X02514). The SalI and XhoI sites were lost as a result of the joining of their compatible sticky ends. The insert can be removed with BspEI or PvuII.

The insert region of 1190 bp and the short flanking regions to the vector junctions were sequenced twice in each direction in order to establish the plasmid sequence. In addition, a total of 63 analytical restriction digests and one minus-enzyme control were performed as detailed in the following table:

		Expected band sizes						
Digest	Well	bp	bp	bp	bp	bp	bp	bp
Uncut	1	1190						
BamHI	2	140	1050					
BsrGI	3	290	900					
HindIII	4	450	740					
NcoI	5	620	570					
SpeI	6	800	390					
AflII	7	990	200					
BamHI BsrGI	8	140	150	900				
BamHI HindIII	9	140	310	740				
BamHI NcoI	10	140	480	570				
BamHI SpeI	11	140	660	390				
BamHI AflII	12	140	850	200				
BsrGI HindIII	13	290	160	740				
BsrGI NcoI	14	290	330	570				
BsrGI SpeI	15	290	510	390				
BsrGI AflII	16	290	700	200				
HindIII NcoI	17	450	170	570				
HindIII SpeI	18	450	350	390				
HindIII AflII	19	450	540	200				
NcoI SpeI	20	620	180	390				
NcoI AflII	21	620	370	200				
SpeI AflII	22	800	190	200				
BamHI BsrGI HindIII	23	140	150	160	740			
BamHI BsrGI NcoI	24	140	150	330	570			
BamHI BsrGI SpeI	25	140	150	510	390			
BamHI BsrGI AflII	26	140	150	700	200			

BamHI HindIII NcoI	27	140	310	170	570			
BamHI HindIII SpeI	28	140	310	350	390			
BamHI HindIII AflII	29	140	310	540	200			
BamHI NcoI SpeI	30	140	480	180	390			
BamHI NcoI AflII	31	140	480	370	200			
BamHI SpeI AflII	32	140	660	190	200			
BsrGI HindIII NcoI	33	290	160	170	570			
BsrGI HindIII SpeI	34	290	160	350	390			
BsrGI HindIII AflII	35	290	160	540	200			
BsrGI NcoI SpeI	36	290	330	180	390			
BsrGI NcoI AflII	37	290	330	370	200			
BsrGI SpeI AflII	38	290	510	190	200			
HindIII NcoI SpeI	39	450	170	180	390			
HindIII NcoI AflII	40	450	170	370	200			
HindIII SpeI AflII	41	450	350	190	200			
NcoI SpeI AflII	42	620	180	190	200			
BamHI BsrGI HindIII NcoI	43	140	150	160	170	570		
BamHI BsrGI HindIII SpeI	44	140	150	160	350	390		
BamHI BsrGI HindIII AflII	45	140	150	160	540	200		
BamHI BsrGI NcoI SpeI	46	140	150	330	180	390		
BamHI BsrGI NcoI AflII	47	140	150	330	370	200		
BamHI BsrGI SpeI AflII	48	140	150	510	190	200		
BamHI HindIII NcoI SpeI	49	140	310	170	180	390		
BamHI HindIII NcoI AflII	50	140	310	170	370	200		
BamHI HindIII SpeI AflII	51	140	310	350	190	200		
BamHI NcoI SpeI AflII	52	140	480	180	190	200		
BsrGI HindIII NcoI SpeI	53	290	160	170	180	390		
BsrGI HindIII NcoI AflII	54	290	160	170	370	200		
BsrGI HindIII SpeI AflII	55	290	160	350	190	200		
BsrGI NcoI SpeI AflII	56	290	330	180	190	200		
HindIII NcoI SpeI AflII	57	450	170	180	190	200		
BamHI BsrGI HindIII NcoI SpeI	58	140	150	160	170	180	390	
BamHI BsrGI HindIII NcoI AflII	59	140	150	160	170	370	200	
BamHI BsrGI HindIII SpeI AflII	60	140	150	160	350	190	200	
BamHI BsrGI NcoI SpeI AflII	61	140	150	330	180	190	200	
BamHI HindIII NcoI SpeI AflII	62	140	310	170	180	190	200	
BsrGI HindIII NcoI SpeI AflII	63	290	160	170	180	190	200	
BamHI BsrGI HindIII NcoI SpeI AflII	64	140	150	160	170	180	190	200

All of these digests produced the expected fragment patterns on agarose gel electrophoresis and a number of these, shaded in the table above, are shown in figure 1. The restriction digests illustrated in figure 1

were carried out using the following conditions:

Digest 1 (Minus enzyme control)

10 µl	100 µg/ml internal control plasmid DNA in 1x NEB buffer #2 + 100 µg/ml BSA
90 µl	1x NEB buffer #2 + 100 µg/ml BSA

Digest 5

15 µl	0.33 U/µl *NcoI* in 1x NEB buffer #2 + 100 µg/ml BSA
10 µl	100 µg/ml internal control plasmid DNA in 1x NEB buffer #2 + 100 µg/ml BSA
75 µl	1x NEB buffer #2 + 100 µg/ml BSA

Digest 9

15 µl	0.33 U/µl *BamHI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *HindIII* in 1x NEB buffer #2 + 100 µg/ml BSA
10 µl	100 µg/ml internal control plasmid DNA in 1x NEB buffer #2 + 100 µg/ml BSA
60 µl	1x NEB buffer #2 + 100 µg/ml BSA

Digest 37

15 µl	0.33 U/µl *BsrGI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *NcoI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *AflII* in 1x NEB buffer #2 + 100 µg/ml BSA
10 µl	100 µg/ml internal control plasmid DNA in 1x NEB buffer #2 + 100 µg/ml BSA
45 µl	1x NEB buffer #2 + 100 µg/ml BSA

Digest 49

15 µl	0.33 U/µl *BamHI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *HindIII* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *NcoI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *SpeI* in 1x NEB buffer #2 + 100 µg/ml BSA
10 µl	100 µg/ml internal control plasmid DNA in 1x NEB buffer #2 + 100 µg/ml BSA
30 µl	1x NEB buffer #2 + 100 µg/ml BSA

Digest 61

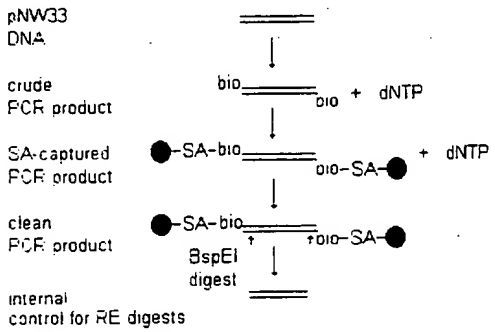
15 µl	0.33 U/µl *BamHI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *BsrGI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *NcoI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *SpeI* in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	0.33 U/µl *AflII* in 1x NEB buffer #2 + 100 µg/ml BSA
10 µl	100 µg/ml internal control plasmid DNA in 1x NEB buffer #2 + 100 µg/ml BSA
15 µl	1x NEB buffer #2 + 100 µg/ml BSA

All 64 restriction digests were incubated at 37°C for 6 hours and samples were then electrophoresed on 2.5 % FMC MetaPhor agarose gels as illustrated in figure 1. The lanes marked M contain mixed samples from all of the digests 1-64 and these lanes were used as size markers after confirmation of the fragment sizes against Stratagene Kb ladder markers.

When spiked into a genomic DNA digest the internal control restriction fragment pattern produced is indicative of both the degree of digestion and the nature of any partial restriction at less than limit digestion. It is possible to deduce from the bands present which, if any, of the enzymes have failed to cut and therefore to take action to correct this before the DNA is used in a subsequent analysis or enrichment procedure.

Preparation of *BspEI*-released internal control DNA on a large scale

The method used for generating internal control PCR product DNA (free from contaminating dNTPs) is depicted in the following figure:



Primers and PCR

20 µM BIO140UP

5' biotin-CGCAGCTGGTAATCCGGACGCCCGCGTCGAAGATGTT 3'

20 µM BIO200DOWN

5' biotin-CGCAGCTGGTAATCCGGACCCGCCGCCGTTGTTGTT 3'

Bulk PCR amplification (192x 100 µl reactions) was carried out according to the conditions described below:

	per 20 ml	final conditions
100 mM dATP	20 µl	100 µM
100 mM dCTP	20 µl	100 µM
100 mM dGTP	20 µl	100 µM
100 mM dTTP	20 µl	100 µM
10x PCR buffer	2 ml	1x
25 mM $MgCl_2$	400 µl	1/50th volume
1 ng/µl pNW33	400 µl	2 ng per PCR
20 µM BIO140UP	200 µl	20 pmol per PCR
20 µM BIO200DOWN	200 µl	20 pmol per PCR
Taq DNA polymerase	100 µl	2.5 U per PCR

The master mix was rapidly dispensed into PCR tubes. Thermal cycling was initiated using the following parameters: 30 cycles of 97°C for 1 min, 50°C for 2 min, and 72°C for 3 min; 72°C for 5 min; and then 4°C.

After PCR amplification, samples were pooled and subjected to capture of biotinylated PCR product termini and BspEI release of internal control DNA.

Capture of biotinylated PCR product termini and *BspEI* release of internal control DNA

All separations were carried out using a Dynal MPC-1 separator (Dynal, product #12001).

20 ml of pooled PCR reaction were mixed with 20 ml of Dynabeads M-280 in 20 mM tris-HCl (pH 7.4), 2 mM EDTA (pH 8.0), 2 M NaCl. The tube was incubated at room temperature for 1 hour with rolling.

The Dynabeads were then washed four times with 20 ml of 10 mM tris-HCl (pH 7.4), 1 mM EDTA (pH 8.0), 1 M NaCl.

BspEI digestion was carried out for 1 hour at 37°C in 2 ml of 0.5 U/µl BspEI in 1x NEB buffer #3. The digest was then ethanol precipitated by the addition of 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ethanol, chilling to -20°C and then centrifugation. Pellets were rinsed with 70 % ethanol prior to redissolution in 500 µl of 1x TE buffer.

1 µl of the BspEI-released internal control DNA was mixed with 10 µl of 50 % glycerol AGE loading dye and electrophoresed on a 1.5 % agarose gel to confirm that the size of the purified DNA was in accordance with that expected.

High molecular weight genomic DNA digestion in the presence of internal control DNA with a dilution series of a mixture of the six set 1 6 bp cutters

High molecular weight canine genomic DNA was first mixed with a dilution series of a mixture of the six set 1 6 bp cutters – each at the same number of units. Aliquots were then removed and mixed with the BspEI-released internal control DNA described above. 20 µg of canine genomic DNA was digested with 0.25, 0.025, 0.0025, 0.00025, 0.000025, 0.0000025 and 0 U/µl BamH1 / BsrGI / HindIII / NcoI / SpeI / AflII in 200 µl. 0.4 µg of canine genomic DNA and 1 µl of BspEI-released internal control DNA were digested with 0.25, 0.025, 0.0025, 0.00025, 0.000025, 0.0000025 and 0 U/µl BamH1 / BsrGI / HindIII / NcoI / SpeI / AflII in 4 µl. Vistra Green staining was used to monitor the extent of internal control DNA digestion and high molecular weight DNA digestion. The fragment pattern from the internal control DNA is diagnostic of both the degree of digestion and the nature of any partial restriction at less than limit digestion.

A premix of restriction enzymes, buffer and BSA was prepared as detailed below:

component	µl
100 U/µl *BamHI*	1 µl
10 U/µl *BsrGI*	10 µl
40 U/µl *HindIII*	2.5 µl
50 U/µl *NcoI*	2 µl
50 U/µl *SpeI*	2 µl
20 U/µl *AflII*	5 µl
10x NEB buffer #2	4 µl
100x BSA	0.4 µl
water	13. 1 µl

Total volume = 40 µl

The premix therefore contained each restriction enzyme at

2.5 U/µl in 1x NEB buffer #2 and 1x BSA. Serial 10-fold dilutions of this premix were then prepared in 1x NEB buffer #2 and 1x BSA.

Premixes of canine genomic DNA, buffer and BSA were also prepared as detailed below:

Component	µl
1 mg/ml canine genomic DNA	20 µl
10x NEB buffer #2	8 µl
100x BSA	0.8 µl
Water	51. 2 µl

These premixes therefore contain canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA.

Canine genomic DNA and restriction enzyme mixes were then set up as follows:

tube	component #1	component #2
1	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of each restriction enzyme at 2.5 U/µl in 1x NEB buffer #2 and 1x BSA
2	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of each restriction enzyme at 0.25 U/µl in 1x NEB buffer #2 and 1x BSA
3	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of each restriction enzyme at 0.025 U/µl in 1x NEB buffer #2 and 1x BSA
4	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of each restriction enzyme at 0.0025 U/µl in 1x NEB buffer #2 and 1x BSA
5	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of each restriction enzyme at 0.00025 U/µl in 1x NEB buffer #2 and 1x BSA
6	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of each restriction enzyme at 0.000025 U/µl in 1x NEB buffer #2 and 1x BSA
7	80 µl of canine genomic DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	20 µl of 1x NEB buffer #2 and 1x BSA

2 µl aliquots were then removed from tubes 1-7 and added to 1 µl of BspEI-released internal control DNA and 1 µl of 2x NEB buffer #2 and 2x BSA to give samples 1-7ic. Samples 1-7ic were then overlaid with 50 µl of mineral oil in order to prevent evaporation.

The remaining volume from tubes 1-7 was then added to 100 µl of 1x NEB buffer #2 and 1x BSA.

All samples were finally incubated at 37°C overnight.

After digestion, 20 µl of digests 1-7 were mixed with 10 µl of 50 % glycerol AGE loading dye and 4 µl of digests 1-7ic were mixed with 2 µl of 50 % glycerol AGE loading dye. Digests in loading dye were then electrophoresed on a 2.5 % MetaPhor agarose gel in 1x TBE. The gel was stained for 60 min in 500 ml of 1x TBE containing 50 µl of Vistra Green. The stained gel was finally imaged on a Fluorimager with the following settings: a 488 nm laser; a 570 DF 30 filter; a PMT setting of 700 V; 200 µm resolution; and low sensitivity.

Restriction digests - final conditions

Restriction digests 1-7 were therefore performed under the following conditions:

tube	µg canine genomic DNA	U *BamH1*	U *BsrGI*	U *HindIII*	U *NcoI*	U *SpeI*	U *AflII*	NEB buffer #2	µg/ml BSA
1	20	50	50	50	50	50	50	1x	100
2	20	5	5	5	5	5	5	1x	100
3	20	0.5	0.5	0.5	0.5	0.5	0.5	1x	100
4	20	0.05	0.05	0.05	0.05	0.05	0.05	1x	100
5	20	0.005	0.005	0.005	0.005	0.005	0.005	1x	100
6	20	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	1x	100
7	20	0	0	0	0	0	0	1x	100

Total volume = 200 µl

Likewise, restriction digests 1-7ic were performed under the following conditions:

tube	µg canine genomic DNA	µl IC DNA	U *BamH1*	U *BsrGI*	U *HindIII*	U *NcoI*	U *SpeI*	U *AflII*	NEB buffer #2	µg/ml BSA
1ic	0.4	1	1	1	1	1	1	1	1x	100
2ic	0.4	1	0.1	0.1	0.1	0.1	0.1	0.1	1x	100
3ic	0.4	1	0.01	0.01	0.01	0.01	0.01	0.01	1x	100
4ic	0.4	1	0.001	0.001	0.001	0.001	0.001	0.001	1x	100
5ic	0.4	1	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	1x	100
6ic	0.4	1	0.00001	0.00001	0.00001	0.00001	0.00001	0.00001	1x	100
7ic	0.4	1	0	0	0	0	0	0	1x	100

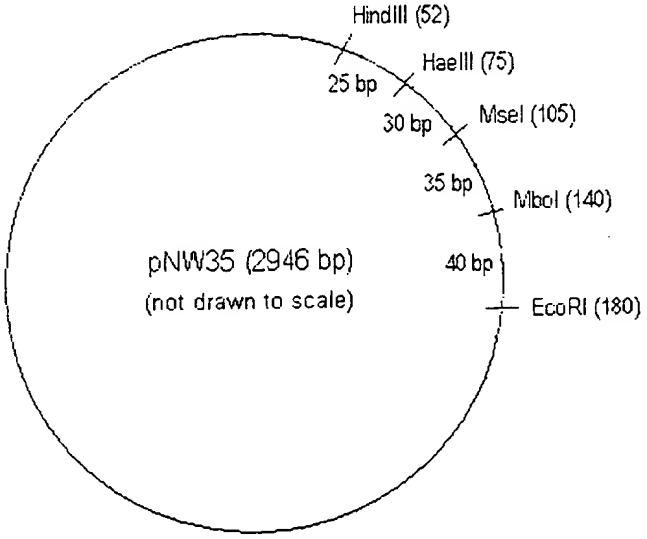
IC = BspEI-released internal control DNA

Total volume = 4 µl

The results are shown in figure 2.

Example 1b – The digestion internal control plasmid for the 4 bp cutter set 1 TRSPA enzymes *HaeIII, MboI,* and *MseI*

The plasmid pNW35 (shown below) was constructed to contain an insert with all of the 4 bp cutter TRSPA enzyme sites.



HindIII and EcoRI sites define the outer ends of the 25 bp and the 40 bp fragments. The sequence of pNW35 is shown below with the inserted region shown in bold type:

atgaccatgattacgccaagctctaatacgactcactatagggaa**agcttccggacgtctcaggctaatgttggcccacc**
gacgttccacgatggggcgctcttaagggcttagaccctcgtcgggagtatttctgtgatctggcgacactcacgcg
agaagtcattaccggcgatatgaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctggcgttacccaactta
atcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcg
cagcctgaatggcgaatgggaaattgtaaacgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttttaacca
ataggccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaag
agtccactattaaagaacgtggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacgtgaacc
atcaccctaatcaagtttttggggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccgatttagagcttg
acggggaaagccggcgaacgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggcaa
gtgtagcggtcacgctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcaggtggcacttttcg
gggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaat
gcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgccttcctgttt
ttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactggatctc
aacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggt

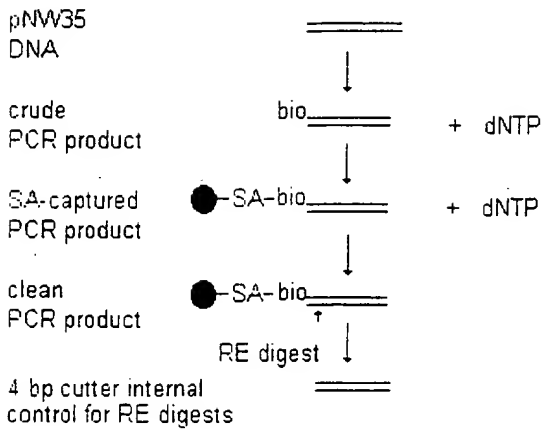
attatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtc
acagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatgagtgataacactgcggcc
aacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaacatgggggatcatgtaactcgccttgat
cgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggcaacaacgtt
gcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcag
gaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcatt
gcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagtcaggcaactatggatgaacga
aatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatatatactttagattgat
ttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttcc
actgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaa
aaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcg
cagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctc
tgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataa
ggcgcagcggtcgggctgaacgggggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatac
ctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcg
gaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgactt
gagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggcc
ttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgc
tcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctc
tccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaat
taatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaattgtgagcggataa
caatttcacacaggaaacagct

The 4 bp internal control plasmid for restriction enzymes HaeIII, MboI, and MseII was prepared by the insertion of a synthetic 130 bp fragment into HindIII / EcoRI-digested pMOSblue (Amersham Pharmacia Biotech).

The insert region of 130 bp was sequenced twice in each direction in order to establish the plasmid sequence. The presence of the restriction sites and the mobility of the fragments released were also checked by restriction digestion and polyacrylamide gel electrophoresis.

Preparation of the internal control spike DNA from the plasmid

The following illustration describes the strategy used for the preparation of EcoRI-released internal control DNA:



Primers and PCR

U-19 mer bio primer	5' bio-GTTTTCCCAGTCACGACGT 3'
ICPCR(F) primer	5' TCCGGACGTCTCAGGCTAATGTT 3'

Bulk PCR amplification (96x 100 µl reactions, repeated to give 192 reactions in total) was carried out according to the conditions described below (all volumes are in µl):

	x1 (µl)	x100 (µl)	Final conditions
Water	84.4	8440	-
10x PCR buffer	10	1000	1x
25 mM dNTP mix	0.8	80	200 µM
20 µM T7 promoter-bio primer	1	100	20 pmol per PCR
20 µM U19 mer-bio primer	1	100	20 pmol per PCR
1 ng/µl pNW35	2	200	2 ng per PCR
5 U/µl *Taq* polymerase	0.8	80	4U per PCR
Total	**100**	**10000**	

The PCR master mix was rapidly dispensed into 96 PCR tubes. Thermal cycling was initiated using the following parameters: 94°C for 2 min; 50 °C for 2 min; 29 cycles of 72°C for 2 min, 94°C for 45 sec, and 50°C for 1 min; 72°C for 8 min; and then 4°C.

Capture of biotinylated PCR product termini and *EcoRI*- release of internal control DNA

All separations were carried out using a Dynal MPC-1 separator (Dynal, product #12001).

20 ml of pooled PCR reaction were mixed with 20 ml of Dynabeads M-280 in 20 mM tris-HCl (pH 7.4), 2 mM EDTA (pH 8.0), 2 M NaCl. The tube was incubated at room temperature for 1 hour with mixing on a Denley Spiramix 5.

The Dynabeads were then washed four times with 20 ml of 10 mM tris-HCl (pH 7.4), 1 mM EDTA (pH 8.0), 1 M NaCl. A fifth wash was performed in 20 ml of 1x buffer M.

EcoRI-digestion was carried out for 1 hour at 37°C in 5 ml of 0.25 U/µl EcoRI in 1x buffer M. The digest was then divided into ten 500 µl aliquot parts. Each aliquot was ethanol precipitated by the addition of 1 µl of See DNA, 0.1 volume of 3 M sodium acetate (pH 5.2), and 2.5 volumes of ethanol. The precipitations were mixed and chilled to 0°C on ice for 30 minutes and then centrifugation at 20,000 maxRCF for 10 minutes. The pellets were rinsed with 70 % ethanol before dissolving in a total of 500 µl of 1x TE buffer.

A 1 µl aliquot of the EcoRI-released internal control DNA was electrophoresed on an 8 % polyacrylamide gel to confirm that the size of the purified DNA was in accordance with that expected.

High molecular weight genomic DNA digestion in the presence of the 4 bp cutter internal control DNA with a dilution series of the enzymes *HaeIII, MseI,* and *MboI*

High molecular weight human placental DNA was mixed with a dilution series of a mixture of the three set 1 4 bp cutter restriction endonucleases – each at the same number of units. Aliquots were then removed and mixed with the EcoRI-released internal control DNA. 3.6 µg of Human placental DNA (Sigma) was digested with 0.5 U/µl, 0.1 U/µl, 0.02 U/µl, 0.004 U/µl, and 0 U/µl each of HaeIII, MboI, and MseI in a total volume of 36 µl. 0.4 µg of placental DNA and 1 µl of EcoRI-released internal control DNA were digested with 0.5 U/µl, 0.1 U/µl, 0.02 U/µl, 0.004 U/µl, and 0 U/µl each of HaeIII, MboI, and MseI in a total volume of 4 µl. Vistra Green (Amersham Pharmacia Biotech) staining was used to monitor the extent of internal control DNA digestion and high molecular weight DNA digestion.

A premix of restriction enzymes (5 U/µl each enzyme), buffer, and BSA was prepared as described below:

Component	µl
50 U/µl *HaeIII*	1
20 U/µl *MseI*	2.5
25 U/µl *MboI*	2
10x NEB buffer #2	1
10x BSA	1
Water	2.5
Total	**10**

Serial 5-fold dilutions of the premix were prepared in 1x NEB buffer #2 and 1x BSA.

A 6x premix of human placental DNA, buffer, and BSA were also prepared as described below:

Component	6x mix (µl)	per reaction (µl)
1mg/ml human placental DNA	24	4
10x NEB buffer #2	9.6	1.6
10x BSA	9.6	1.6
Water	52.8	8.8

This premix contained placental DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA.

Human placental DNA and restriction mixes were then set up as follows:

Tube	Component #1	Component #2
1	16 µl of placental DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	4 µl of each restriction enzyme at 5 U/µl in 1x NEB buffer #2 and 1x BSA
2	16 µl of placental DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	4 µl of each restriction enzyme at 1 U/µl in 1x NEB buffer #2 and 1x BSA
3	16 µl of placental DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	4 µl of each restriction enzyme at 0.2 U/µl in 1x NEB buffer #2 and 1x BSA
4	16 µl of placental DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	4 µl of each restriction enzyme at 0.04 U/µl in 1x NEB buffer #2 and 1x BSA
5	16 µl of placental DNA at 0.25 mg/ml in 1x NEB buffer #2 and 1x BSA	4 µl of 1x NEB buffer #2 and 1x BSA

2 µl aliquots were then removed from tubes 1-5 and added to 1 µl of EcoRI-released internal control DNA and 1 µl of 2x NEB buffer #2 and 2x BSA to give samples 1-5ic. Samples 1-5ic were then overlaid with 50 µl of mineral oil in order to prevent evaporation.

The remaining volume from tubes 1-5 was then added to 18 µl of 1x NEB buffer #2 and 1x BSA.

All samples were finally incubated at 37°C overnight.

Restriction digests - final conditions

Restriction digests 1-5 were therefore performed under the following conditions:

tube	µg human placental DNA	U *HaeIII*	U *MseI*	U *MboI*	NEB buffer #2	µg/ml BSA
1	3.6	20	20	20	1x	100
2	3.6	4	4	4	1x	100
3	3.6	0.8	0.8	0.8	1x	100
4	3.6	0.16	0.16	0.16	1x	100
5	3.6	0	0	0	1x	100

Total volume = 36 µl

Likewise, restriction digests 1-5ic were performed under the following conditions:

tube	µg human placental DNA	µl IC DNA	U *HaeIII*	U *MseI*	U *MboI*	NEB buffer #2	µg/ml BSA
1	0.4	1	2	2	2	1x	100
2	0.4	1	0.4	0.4	0.4	1x	100
3	0.4	1	0.08	0.08	0.08	1x	100
4	0.4	1	0.016	0.016	0.016	1x	100
5	0.4	1	0	0	0	1x	100

IC = *EcoRI*-released internal control DNA

Total volume = 4 µl

10 µl of digests 1-5 were each mixed with 3 µl of loading dye and 4 µl of digests 1-5ic were mixed with 1 µl of loading dye. To sample number 5ic, 180 ng of PCR molecular weight markers were added to serve as size standards. The band sizes for these markers are 50 bp, 100 bp, 200 bp, 300 bp, 400 bp, 500 bp, 525 bp, 700 bp, and 1000 bp. The digests

were then electrophoresed on an 8 % polyacrylamide gel in 1x TBE as described below:

10 ml of 10x TBE
19.5 ml of 40 % acrylamide
19 ml of 2 % methylene bisacrylamide
50.4 ml of water
1 ml of freshly prepared 10 % (w/v) ammonium persulphate
100 µl of TEMED

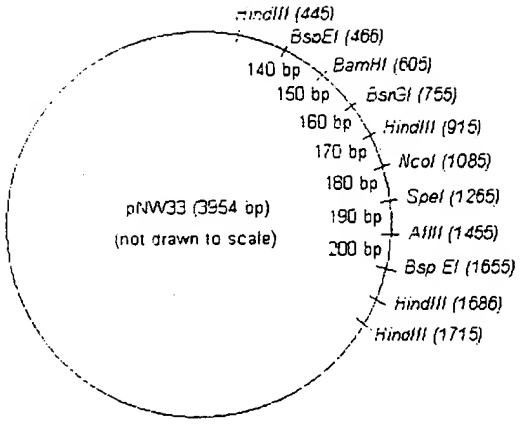
A Cambridge Electrophoresis Ltd. vertical protein electrophoresis unit was used with 1 mm plate spacing. The samples were electrophoresed at 30 mA for 2 hours in 1x TBE. The gel was then stained for 30 minutes in 500 ml of 1x TBE containing 50 µl of Vistra Green. The stained gel was imaged on a Fluorimager with the following settings: a 488 nm laser; a 570 DF 30 filter; a PMT setting of 700 V; 200 µm resolution; and low sensitivity.

The results are shown in figure 3.

Example 2

TRSPA-2 analysis with the pNW33 *BamHI*, *HindIII*, and *AflII* matrix and the pNW33 *HindIII*, *NcoI*, and *SpeI* matrix probed with the 140 bp *BspEI* to *BamHI* fragment from pNW33

As an example of the case where the probe hybridizes to an arrayed PCR fragment with a different restriction site at each end, the pNW33 BamHI, HindIII, and AflII matrix (matrix #7) was probed with a PCR product from the 140 bp BspEI (466) to BamHI (605) restriction fragment within pNW33. In this example, the probe binds to a 204 bp HindIII to BamHI PCR product derived from the 160 bp HindIII (445) to BamHI (605) restriction fragment within pNW33 (see below).



As an example of the case where the probe hybridizes to an arrayed PCR fragment with the same restriction site at each end, the pNW33 HindIII, NcoI, and SpeI matrix (matrix #17) was probed with a PCR product from the 140 bp BspEI (466) to BamHI (605) restriction fragment within pNW33. In this example, the probe binds to a 514 bp HindIII to HindIII PCR product derived from the 470 bp HindIII (445) to HindIII (915) restriction fragment within pNW33 (see above).

Oligonucleotides

BamHI short PCR primer	5' TGTAACGACACATTGCTGGATACC 3'
HindIII short PCR primer	5' ATATAACTCTCGCTCCTTGATAAC 3'
NcoI short PCR primer	5' AGGCGTCTGAGGCTGCGGCTATGG 3'
SpeI short PCR primer	5' AACCCGTCGCGACGAGAGTCTAAG 3'
AflII short PCR primer	5' GATATACGTGATATATTTTGATTG 3'
BamHI adaptor	5' pGATCGGTATCCAGCAATGTGTCGTTACA 3'
HindIII adaptor	5' pAGCTGTTATCAAGGAGCGAGAGTTATAT 3'
NcoI adaptor	5' pCATGCCATAGCCGCAGCCTCAGACGCCT 3'
SpeI adaptor	5' pCTAGCTTAGACTCTCGTCGCGACGGGTT 3'

AflII adaptor	5' pTTAACAATCAAAATATATCACGTATATC 3'
BamHI long PCR primer	5' TGTAACGACACATTGCTGGATACCGATCC 3'
HindIII long PCR primer	5' ATATAACTCTCGCTCCTTGATAACAGCTT 3'
NcoI long PCR primer	5' AGGCGTCTGAGGCTGCGGCTATGGCATGG 3'
SpeI long PCR primer	5' AACCCGTCGCGACGAGAGTCTAAGCTAGT 3'
AflII long PCR primer	5' GATATACGTGATATATTTTGATTGTTAAG 3'

Luc140down primer

5' GCGCTAGGGATCCTTACTGGGACGAAGACGAA 3'

Luc140up-bio primer

5' biotin-CGCAGCTGGTAATCCGGACGCCCGCGTCGAAGATGTT3'

Restriction digestion of pNW33

Restriction digests of pNW33 were set up by combining the following:

Matrix #7

450 µl	22 µg/ml pNW33 in 1x NEB buffer #2 + 100 µg/ml BSA
50 µl	0.6 U/µl *BamHI* in 1x NEB buffer #2 + 100 µg/ml BSA
50 µl	0.6 U/µl *HindIII* in 1x NEB buffer #2 + 100 µg/ml BSA
50 µl	0.6 U/µl *AflII* in 1x NEB buffer #2 + 100 µg/ml BSA

Matrix #17

450 µl	22 µg/ml pNW33 in 1x NEB buffer #2 + 100 µg/ml BSA
50 µl	0.6 U/µl *HindIII* in 1x NEB buffer #2 + 100 µg/ml BSA
50 µl	0.6 U/µl *NcoI* in 1x NEB buffer #2 + 100 µg/ml BSA
50 µl	0.6 U/µl *SpeI* in 1x NEB buffer #2 + 100 µg/ml BSA

The samples were then vortex mixed, and incubated at 37°C overnight.

Calf intestinal alkaline phosphatase (CIAP) digestion of pNW33 restriction digests

400 µl fractions of the 20 restriction digests (each containing 6.6 µg of digested pNW33) were ethanol precipitated by the addition of 1 µl of See DNA, 0.1 volume of 3 M sodium acetate (pH 5.2), and 2.5 volumes of ethanol, chilling to -20°C, and then centrifugation. Pellets were rinsed with 70 % ethanol prior to re-dissolution in 20 µl of 1x CIAP buffer containing 40 U of CIAP. The CIAP digests were carried out for 5 hours at 37°C and were then made up to 400 µl with 1x TE buffer.

Phenol extraction of pNW33 CIAP digests

The diluted digests were extracted with 400 µl of phenol and then ethanol precipitated as described above but with a 100 % ethanol wash after the 70 % ethanol wash. Samples were finally re-dissolved in 20 µl of TE buffer.

Annealing of short PCR primers to cognate adaptors

Short PCR primers and their cognate adaptors were annealed by adding 1 µl of 200 µM short PCR primer to 1 µl of 200 µM cognate adaptor in 20 µl of 50 mM NaCl, 1x TE buffer. The mixed oligonucleotides were overlaid with 30 µl of light mineral oil and were then heated to 90°C for 5 minutes followed by slow cooling to room temperature. The annealed short PCR primer / cognate adaptor complexes were then diluted with 1 ml of 1x TE buffer and stored frozen at -20°C.

Ligation to annealed short PCR primers and cognate adaptors

1 µl of phenol extracted pNW33 digest was used per ligation. Ligations were performed in 100 µl of 1x ligase buffer containing 1 mM ATP

and 10 µl (100 U) of T4 DNA ligase. 20 µl aliquots from the 1 ml of annealed short PCR primer / cognate adaptor complexes were added according to the following table.

#7	*BamHI / HindIII / AflII*
#17	*HindIII / NcoI / SpeI*

Ligation reactions were carried out for 24 hours at 16°C. Samples were then diluted to 1 ml with TE buffer and stored at -20°C.

The diluted ligation reactions were then further diluted 1 in 10 and 10 µl was used as PCR template per 100 µl reaction.

Restriction digestion of human placental DNA

1 U (~50 µg) of human placental DNA was digested overnight at 37°C with 100 U each of *BamHI*, *BsrGI*, *HindIII*, *NcoI*, *SpeI*, and *AflII* in 400 µl of 1x NEB buffer #2 containing 100 µg/ml BSA.

CIAP digestion of human placental DNA restriction digest

The digest was ethanol precipitated by the addition of 1 µl of See DNA, 0.1 volume of 3 M sodium acetate (pH 5.2), and 2.5 volumes of ethanol, chilling to -20°C, and then centrifugation. The pellet was rinsed with 70 % ethanol prior to re-dissolution in 50 µl of 1x CIAP buffer containing 100 U of CIAP. The CIAP digest was carried out for 5 hours at 37°C and was then made up to 400 µl with 1x TE buffer.

Phenol extraction of human placental DNA CIAP digest

The diluted CIAP digest was extracted with 400 µl of phenol and then ethanol precipitated as described above – again with a 100 % ethanol wash after the 70 % ethanol wash. The sample was finally re-dissolved in 10 µl of TE buffer.

Ligation to annealed short PCR primers and cognate adaptors

Short PCR primers were annealed to their cognate adaptors as described above.

The ligation to annealed short PCR primers and cognate adaptors was carried out in 100 µl of 1x ligase buffer with 1 mM ATP, 100 U of T4 DNA ligase and 10 µl of each of the six short PCR primer / cognate adaptor complexes as above.

The ligation reaction was carried out for 24 hours at 16°C. The sample was then diluted to 1 ml with 1x TE buffer and stored at -20°C.

0.2 µl was used as PCR template per 100 µl reaction.

PCR amplification conditions

An initial touch-down reaction was carried out in 50 µl of 1x PCR buffer with all four dNTPs at 200 µM and *Taq* DNA polymerase at 0.05 U/µl. Long PCR primers were used at 400 nM. 10 µl of pNW33 PCR template was used per reaction and 0.2 µl of human placental DNA PCR template was used per reaction. The samples were overlaid with 40 µl of light mineral oil and were touch-down thermocycled as described below:

98°C for 1 min, 72°C for 2 min, 72°C for 5 min
98°C for 1 min, 69°C for 2 min, 72°C for 5 min
98°C for 1 min, 66°C for 2 min, 72°C for 5 min
98°C for 1 min, 63°C for 2 min, 72°C for 5 min
98°C for 1 min, 60°C for 2 min, 72°C for 5 min

The main thermal cycling was then carried out in 100 µl of 1x PCR buffer with all four dNTPs at 200 µM and *Taq* DNA polymerase at 0.05 U/µl. Long PCR primers were used at 400 nM. The samples were subjected to 20 cycles of 98°C for 1 min, 60°C for 2 min, 72°C for 5 min and then 72°C for 10 min followed by chilling at 10°C and recovery from under oil.

Arraying onto nylon membranes

1 µl aliquots from the PCR amplified samples were spotted onto Hybond N+ nylon membranes. The membranes were then transferred to a stack of three sheets of 3MM paper, saturated with 0.4 M NaOH, and incubated for 10 minutes. The NaOH was rinsed away in 2x SSC and the membranes were used directly for the pre-hybridization.

Probe synthesis

A PCR master mix was prepared as follows (all volumes are in µl):

Master Mix Preparation	140
Water	183.6
10x PCR buffer	30
2 mM dATP	15
2 mM dGTP	15
2 mM dTTP	15
2 mM dCTP	3
140 probe template	1
Luc140up-bio	2.5
Luc140down	2.5
Taq DNA polymerase (5 U/µl)	2.4
Total	**270**

Five PCR reactions were performed. For each reaction, 5 µl of 33P-labelled dCTP was added to the PCR tube. The reactions were then made up to 50 µl by the addition of 45 µl of the master mix to each tube. Each reaction was gently mixed.

PCR cycling parameters

94°C	2 min		
50 °C	2 min		
72°C	2 min	}	25 cycles
94°C	45 sec		
50 °C	1 min		
72°C	8 min		
4°C	hold		

After thermal cycling, the five PCR amplifications were pooled. The 250 µl of labelled DNA was then used for the following magnetic bead purification procedure.

Capture of biotinylated PCR product and release of non-biotinylated strands

All separations were carried out using a Dynal MPC-4 separator (Dynal, product #12004).

250 µl of the pooled PCR was mixed with an equal volume of 10 mg/ml streptavidin-coated colloidal Fe3O4 particles in 20 mM tris-HCl (pH 7.4), 2 mM EDTA (pH 8.0), 2 M NaCl. The tube was incubated at room temperature for 1 hour with mixing on a Denley Orbital Mixer.

The streptavidin-coated colloidal Fe3O4 particles were then washed with 1 ml of 10 mM tris-HCl (pH 7.4), 1 mM EDTA (pH 8.0), 1 M NaCl. Three more identical washes were performed.

The washed streptavidin-coated colloidal Fe3O4 particles were incubated in 500 µl of 0.1 M NaOH for 10 minutes at room temperature. The supernatant was removed and added to 500 µl of 0.5 M

HEPES. The samples were then ethanol precipitated by the addition of 0.1 volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ethanol, chilling to 0°C (on ice for 30 minutes), and then centrifugation at 20,000 maxRCF for 10 minutes. The pellets were rinsed with 70 % ethanol before dissolving in 100 µl of TE buffer.

Hybridization

Each membrane was placed in a 55 mm x 35 mm x 21 mm plastic box and 1.25 ml of pre-hybridization solution (5x SSC; Denhardt's solution; 1 % SDS; 10 % dextran sulphate [Mw 500,000]; 0.3 % tetrasodium pyrophosphate; 100 µg/ml denatured, sonicated DNA – pre-warmed to 65°C) was added. Each box was closed and incubated at 65°C for 50 minutes on a rocking platform. The pre-hybridization solution was removed and replaced with hybridization solution (5x SSC; Denhardt's solution; 1 % SDS; 10 % dextran sulphate [Mw 500,000]; 0.3 % tetrasodium pyrophosphate; 100 µg/ml denatured, sonicated DNA – containing 5 µl of the appropriate 33P-labelled probe) and the box was incubated at 65°C for 3 hours on a rocking platform.

Washing

The membranes were drained and transferred to 200 ml of 2x SSC, 0.1 % SDS at 68°C for 30 minutes. A further wash was carried out in 0.2x SSC, 0.1 % SDS at 71°C for 30 minutes. The membranes were rinsed in 2x SSC at room temperature and laid out on blotting paper to remove excess liquid. Once dry, the membrane was covered in Saran Wrap and exposed to a Kodak Phosphor Screen for 1 hour. The phosphor screen was subsequently imaged using a Molecular Dynamics Storm 860 Phosphorimager.

Results are shown in figures 4 and 5.

Example #3a

A single cycle of inter-population perfectly matched duplex depletion wherein E.*coli* MutS protein is used to capture an A/A mismatch-containing duplex

'Affected' versus 'unaffected' (i.e. inter-population) mismatch-containing duplex selection can be achieved by: attaching a mismatch-binding protein to a solid support (or using the mismatch-binding protein in solution followed by subsequent solid-phase capture); taking denatured 'affected' DNA fragments and hybridizing these to denatured and biotinylated 'unaffected' DNA fragments; and capture of mismatch-containing duplex molecules with the mismatch-binding protein. Releasing the mismatch-containing duplex molecules (without strand denaturation), streptavidin capture and then release of the non-biotinylated strands will give only the desired species.

In this example, PCR fragments are prepared and used to demonstrate each of the individual steps for a single cycle of inter-population perfectly matched duplex depletion using E.*coli* MutS protein.

Clone insert design

The clone inserts were constructed using standard cloning methodology well known to those skilled in the art and were inserted between the *AvaI* site and *EcoRI* of pMOS*Blue* (Amersham Pharmacia Biotech).

The clone inserts contain a common 9 base pair internal core sequence in which a single nucleotide change or an insertion can be located. The internal core sequence is derived from codons 272-274 of human p53. These codons (GTG CGT GGT) correspond to a mutational hotspot found in lung and other types of cancer (R273L). For the design of the clone inserts containing a mismatch (only one from a complete series of which is shown below), the nucleotide in position 5 of this core sequence is

modified (GTGCXTGGT).

The internal core sequence is flanked by a random sequence – allowing the independent detection of the clone #1 and the clone #7 insert sequence in a mixed population of clone inserts.

Clone #1

Mutant sequence (#1M)

```
5'CCCGGGGGATCCTCGTTTTATTGGGCCGAGTTTTGGTCCGTAGTGCTTGGTTAGATATGCTTAT
3'GGGCCCCCTAGGAGCAAAATAACCCGGCTCAAAACCAGGCATCACGAACCAATCTATACGAATA

GTTCACAAAATCATCCTTGTACAGAATTC3'
CAAGTGTTTTAGTAGGAACATGTCTTAAG5'
```

Control sequence (#1C)

```
5'CCCGGGGGATCCTCGTTTTATTGGGCCGAGTTTTGGTCCGTAGTGCATGGTTAGATATGCTTAT
3'GGGCCCCCTAGGAGCAAAATAACCCGGCTCAAAACCAGGCATCACGTACCAATCTATACGAATA

GTTCACAAAATCATCCTTGTACAGAATTC3'
CAAGTGTTTTAGTAGGAACATGTCTTAAG5'
```

Clone #7

Control sequence (#7C)

```
5'CCCGGGTGTACACAAAAGTTTACCTGAAGAACGTGGGGGGTCGTGCCTGGTCTTGCGTCACCTG
3'GGGCCCACATGTGTTTTCAAATGGACTTCTTGCACCCCCCAGCACGGACCAGAACGCAGTGGAC

GTCTCAGGAGAGGGTCCCCATGGGAATTC3'
CAGAGTCCTCTCCCAGGGGTACCCTTAAG5'
```

Preparation of upper strand 5'-biotinylated and lower strand 5'-biotinylated #1C double-stranded PCR product, and upper strand 5'-biotinylated and lower strand 5'-biotinylated #7C double-stranded PCR product

Oligonucleotides

BIOUPST2	5' bio-CTACTGATCGGATCCCCG 3'
BIODOWN3	5' bio-AAACGACGGCCAGTGAAT 3'

PCR reaction set-up (#1C)

#1C in pMOS*Blue* at 2.50 µg/ml	100 µl
Water	8400 µl
10x PCR buffer	1000 µl
50 µM BIOUPST2	100 µl
50 µM BIODOWN3	100 µl
10 mM dNTPs	200µl
Taq DNA polymerase (5 U/µl)	100 µl

PCR reaction set-up (#7C)

#7C in pMOS*Blue* at 3.32 µg/ml	100 µl
Water	8400 µl
10x PCR buffer	1000 µl
50 µM BIOUPST2	100 µl
50 µM BIODOWN3	100 µl
10 mM dNTPs	200µl
Taq DNA polymerase (5 U/µl)	100 µl

96x 200 µl reactions were carried out for template #1C and 96x 200 µl reactions were carried out for template #7C on a 96-well Perkin Elmer Cetus GeneAmp PCR System 9600 machine as described below:

95 °C, 5 minutes	1 cycle
95 °C, 1 minute	30 cycles
50 °C, 1 minute	
72 °C, 1 minute	
72 °C, 5 minutes	1 cycle
4°C, hold	

The PCR products were pooled together and precipitated by adding 0.1 volumes of 3 M sodium acetate and 1 volume of isopropanol followed by centrifugation at 16,000 rpm for 30 minutes at 4°C (in a Centrikon T-2070 ultracentrifuge; swinging bucket Kontron rotor TST 41.14). Pellets were washed with 14 ml of ethanol and centrifuged at 20,000 rpm for 30 minutes. Finally, the pellets were air-dried and resuspended in a total volume of 0.6 ml of TE buffer.

The 0.6 ml PCR sample was purified in twelve Microspin S-300 HR columns (50 µl per column) following the manufacturer's protocol. Briefly, the resin in the columns was resuspended by vortexing. Columns were centrifuged at 735 x g (3000 rpm in a microfuge) for 1 minute. The sample was then applied to the centre of the resin, being careful not to disturb the bed. The columns were centrifuged at 735 x g for 2 minutes and the flow-through containing the PCR product was collected. The twelve eluted 50 µl volumes were pooled together (pool 1). Columns were washed with 50 µl of TE buffer and the eluted fractions were loaded onto a fresh S-300 HR column. Product yield and removal efficiency of the PCR primers were analysed on a 1.5 % agarose gel.

Preparation of non-biotinylated #1M single-stranded DNA and non-biotinylated #7C single-stranded DNA

Oligonucleotides

BIOUPST2	5' bio-CTACTGATCGGATCCCCG 3'
DOWN3	5' AAACGACGGCCAGTGAAT 3'

PCR reaction set-up (#1M)

#1M in pMOS*Blue* at 3.56 µg/ml	100 µl
Water	8400 µl
10x PCR buffer	1000 µl
50 µM BIOUPST2	100 µl
50 µM DOWN3	100 µl
10 mM dNTPs	200µl
Taq DNA polymerase (5 U/µl)	100 µl

PCR reaction set-up (#7C)

#7C in pMOS*Blue* at 3.32 µg/ml	100 µl
Water	8400 µl
10x PCR buffer	1000 µl
50 µM BIOUPST2	100 µl
50 µM DOWN3	100 µl
10 mM dNTPs	200µl
Taq DNA polymerase (5 U/µl)	100 µl

48x 200 µl reactions were carried out for template #1M and 48x 200 µl reactions were carried out for template #7C on a 96-well Perkin

Elmer Cetus GeneAmp PCR System 9600 machine as described below:

95 °C, 5 minutes	1 cycle
95 °C, 1 minute	30 cycles
50 °C, 1 minute	
72 °C, 1 minute	
72 °C, 5 minutes	1 cycle
4°C, hold	

Capture of biotinylated PCR product strands and release of non-biotinylated strands for the preparation of non-biotinylated #1M single-stranded DNA and non-biotinylated #7C single-stranded DNA

10 ml of pooled #1M and #7C PCR amplifications were each mixed with an equal volume of 4 mg/ml streptavidin-coated colloidal Fe_3O_4 particles taken up in 20 mM tris-HCl (pH 7.4), 2 mM EDTA (pH 8.0), 2 M NaCl. The tubes were incubated at room temperature for 60 minutes with mixing.

The streptavidin-coated colloidal Fe_3O_4 particles were then washed with 20 ml of 10 mM tris-HCl (pH 7.4), 1 mM EDTA (pH 8.0), 1 M NaCl. Two more identical washes were performed.

The washed streptavidin-coated colloidal Fe_3O_4 particles were finally incubated in 800 µl of 0.1 M NaOH for 10 minutes at room temperature. The supernatants were removed and added to 200 µl of 2 M HEPES (free acid). Samples were quantified by absorbance at 260 nm.

Denaturation and annealing

Denaturation and annealing reactions were prepared by mixing:

A = 400 ng of upper strand 5'-biotinylated and lower strand 5'-biotinylated #1C double-stranded PCR product
B = 400 ng of upper strand 5'-biotinylated and lower strand 5'-biotinylated #7C double-stranded PCR product
C = 200 ng of non-biotinylated #1M single-stranded DNA (the lower of the two #1M strands shown above)
D = 200 ng of non-biotinylated #7C single-stranded DNA (the lower of the two #7C strands shown above)

Reannealing between the upper strand of A and the single-stranded C will therefore give rise to an A/A mismatch-containing duplex, 5'-biotinylated on the upper strand, for clone insert #1.

Reannealing between the upper strand of B and the single-stranded D will therefore give rise to a perfectly matched duplex, 5'-biotinylated on the upper strand, for clone insert #7.

0.1 volumes of 1 M NaOH were then added, followed by incubation at room temperature for 10 minutes. 0.25 volumes of 2 M HEPES (free acid) were finally added followed by incubation at 42°C for 1 hour.

Samples were adjusted to 50 µl and were made 1x in PBS and 1 mg/ml in BSA ready for reaction with MutS protein-coated magnetic beads.

One 50 µl sample (the pre-enrichment control, sample 6) was used directly for capture of biotinylated PCR product strands and release of non-biotinylated strands.

Mismatch capture

20 µl of M_2B_2 MutS protein-coated magnetic particles (Genecheck, lot #20) were added to the annealed DNA above. Samples were incubated for 1 hour at room temperature with shaking.

Samples were then washed twice with 200 µl of ice-cold PBS.

Samples were finally eluted from the magnetic beads for 10 minutes at room temperature in 50 µl of the following:

Sample 1	1 M NaCl
Sample 2	PBS
Sample 3	1 M urea
Sample 4	1% (w/v) SDS
Sample 5	10 mM NaOH

Capture of biotinylated PCR product strands and release of non-biotinylated strands

All separations were carried out using an Amersham magnetic separator (RPN1682, batch #1).

50 µl of the eluates from the MutS protein-coated magnetic beads and the pre-enrichment control (sample 6) were each mixed with an equal volume of 4 mg/ml streptavidin-coated colloidal Fe_3O_4 particles in 20 mM tris-HCl (pH 7.4), 2 mM EDTA (pH 8.0), 2 M NaCl. The tubes were incubated at room temperature for 30 minutes with regular mixing.

The streptavidin-coated colloidal Fe_3O_4 particles were then washed twice with 500 µl of 10 mM tris-HCl (pH 7.4), 1 mM EDTA (pH 8.0), 1 M NaCl at room temperature.

The washed streptavidin-coated colloidal Fe_3O_4 particles were incubated in 10 µl of 0.1 M NaOH for 10 minutes at room temperature. The supernatant was removed and added to 2.5 µl of 2 M HEPES (free acid).

5 µl fractions were finally spotted onto Hybond N+ nylon membranes along with 200 ng, 20 ng, 2 ng, and 200 pg amounts of the following:

C = non-biotinylated #1M single-stranded DNA (the lower of the two #1M strands shown previously)
D = non-biotinylated #7C single-stranded DNA (the lower of the two #7C strands shown previously)

Probe labelling

Oligonucleotides

#1 probe oligo 5' GGCCGAGTTTTGGTCCGTAG 3'

#7 probe oligo 5' GTCTTGCGTCACCTGGTCTCAG 3'

Preparation of probes

#1 probe oligo and #7 probe oligo were radioactively 5' end-labelled using T4 polynucleotide kinase as described below (all volumes are in µl):

	#1 probe	#7 probe
#1 probe oligo (25 pmol)	1.25	-
#7 probe oligo (25 pmol)	-	1.25
10x PNK buffer	2.5	2.5
[γ-^{33}P] ATP >92 TBq/mmol, >2500 Ci/mmol, 370 mBq/ml, 10 mCi/ml, (Amersham Pharmacia Biotech., AH9968, lot #B0006)	12.5	12.5
T4 PNK (10 U/µl) (Amersham International, E70031Y, lot #201226)	2.5	2.5
Water	6.25	6.25

The reactions were incubated at 37°C for 30 minutes and

then heated to 70°C for 5 minutes to denature the enzyme.

MicroSpin G-25 column purification

Two G-25 columns (APB 27-5325-01, lot #9015325011) were resuspended by vortexing, and the bottom closures were snapped off as described in the manufacturer's instructions. A pre-spin of the columns was carried out for 1 minute at 730 maxRCF – 2670 rpm in a Hettich Zeutrifugen EBA 12 benchtop centrifuge – and the eluates were discarded. The 25 µl reactions were applied to each column, and the columns were centrifuged for 2 minutes at 730 maxRCF. The eluates from the second spin were stored and used as probes.

Hybridization

Each membrane was placed in a 55 mm x 35 mm x 21 mm plastic box and 2.5 ml of pre-hybridization solution (5x SSC; Denhardt's solution; 1 % SDS; 10 % dextran sulphate [MW 500,000]; 0.3 % tetrasodium pyrophosphate; 100 µg/ml denatured, sonicated DNA – pre-warmed to 42°C) was added. Each box was closed and incubated at 42°C for 1 hour on a rocking platform. The pre-hybridization solution was removed and replaced with hybridization solution (5x SSC; Denhardt's solution; 1 % SDS; 10 % dextran sulphate [MW 500,000]; 0.3 % tetrasodium pyrophosphate; 100 µg/ml denatured, sonicated DNA – containing 2.5 µl of the appropriate ^{33}P-labelled probe) and the box was incubated at 42°C overnight on a rocking platform.

Washing

The membranes were drained and transferred to 200 ml of 2x SSC, 0.1 % SDS at 42°C for 10 minutes. A further wash was carried out in 0.2x SSC, 0.1 % SDS at 42°C for 10 minutes. The membranes were rinsed in 2x SSC at room temperature and laid out on blotting paper to remove excess liquid. Once dry, the membrane was covered in Saran

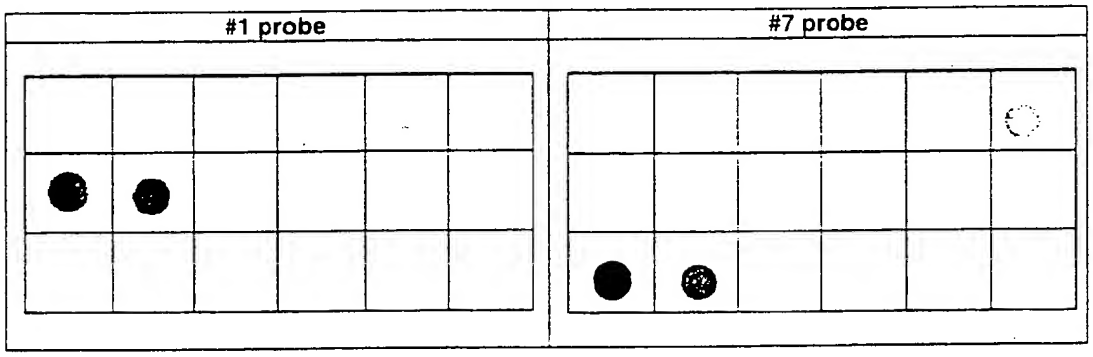
Wrap and exposed to a Kodak Phosphor Screen for 1 hour. The phosphor screen was subsequently imaged using a Molecular Dynamics Storm 860 Phosphorimager.

Dot blot layout

Sample 1 5 µl fraction	Sample 2 5 µl fraction	Sample 3 5 µl fraction	Sample 4 5 µl fraction	Sample 5 5 µl fraction	Sample 6 5 µl fraction
200 ng #1M ssDNA	20 ng #1M ssDNA	2 ng #1M ssDNA	200 pg #1M ssDNA	-	-
200 ng #7C ssDNA	20 ng #7C ssDNA	2 ng #7C ssDNA	200 pg #7C ssDNA	-	-

ssDNA = single-stranded DNA

Probe hybridisation results



Signal intensities from the spots above were reported using ImageQuant 5.0 software (Molecular Dynamics), with the SumAboveBG figures being used after drawing a 6x3 grid over the array of spots.

#1 probe SumAboveBG signals

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
142257	24356	180359	394296	465242	658323
200 ng #1M ssDNA	20 ng #1M ssDNA	2 ng #1M ssDNA	200 pg #1M ssDNA	-	-
15961011	2842937	252392	25316	2281	2475
200 ng #7C ssDNA	20 ng #7C ssDNA	2 ng #7C ssDNA	200 pg #7C ssDNA	-	-
66536	37955	10621	2736	1955	1934

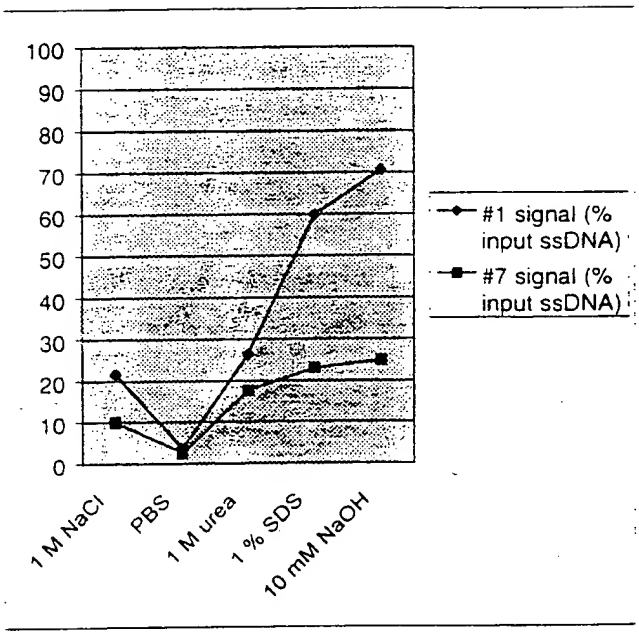
#7 probe SumAboveBG signals

Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6
66392	17071	118109	153372	166211	666853
200 ng #1M ssDNA	20 ng #1M ssDNA	2 ng #1M ssDNA	200 pg #1M ssDNA	-	-
80873	23256	6371	2399	1831	1857
200 ng #7C ssDNA	20 ng #7C ssDNA	2 ng #7C ssDNA	200 pg #7C ssDNA	-	-
10619299	2052027	207827	15470	2055	1844

Recovery

Elution conditions	#1 signal (% input ssDNA)	#7 signal (% input ssDNA)	#1/#7 signal ratio
1 M NaCl	21.6 %	10.0 %	2.2
PBS	3.7 %	2.6 %	1.4
1 M urea	26.4 %	17.7 %	1.5
1 % SDS	59.9 %	23.0 %	2.6
10 mM NaOH	70 .7 %	24.9 %	2.8

Recovery figures are plotted below for the various different elution conditions:



Example #3b

A single cycle of inter-population perfectly matched duplex depletion wherein bacteriophage T4 endonuclease VIII protein containing a cleavage-inactivating N62D point mutation is used to capture an A/A mismatch-containing duplex

In this example, PCR fragments are again prepared and used to demonstrate each of the individual steps for a single cycle of inter-population perfectly matched duplex depletion using bacteriophage T4 endonuclease VIII protein containing a cleavage-inactivating N62D point mutation.

Clone design and DNA preparation

Clone insert design was exactly as described in example #3a (see clone #1: mutant sequence (#1M), control sequence (#1C) and clone #7: control sequence (#7C)).

Preparation of upper strand 5'-biotinylated and lower strand 5'-biotinylated #1C double-stranded PCR product, and upper strand 5'-biotinylated and lower strand 5'-biotinylated #7C double-stranded PCR product were as described in example #3a.

Preparation of non-biotinylated #1M single-stranded DNA and non-biotinylated #7C single-stranded DNA were again as described in example #3a.

Denaturation and annealing

Denaturation and annealing reactions were prepared as described in example #3a.

Reannealing will therefore again give rise to an A/A mismatch-containing duplex, 5'-biotinylated on the upper strand, for clone insert #1 and a perfectly matched duplex, 5'-biotinylated on the upper strand, for clone insert #7.

One 50 µl sample (the pre-enrichment control) was adjusted to 100 µl with TE buffer and was used directly for capture of biotinylated PCR product strands and release of non-biotinylated strands.

Mismatch capture

50 µl samples of the annealing reaction were mixed with an equal volume of 200 mM sodium phosphate (pH 6.5), 100 mM KCl. 10 µg of GST-tagged T4 endonuclease VII N62D mutant (obtained from Prof. Börries Kemper, Univ. Cologne) were added to the annealed DNA and the samples were incubated for 15 minutes at 16°C.

Samples were then mixed with 200 µl of a 50 % slurry of Glutathione Sepharose 4B (Amersham Pharmacia Biotech, lot #279991) in 100 mM sodium phosphate (pH 6.5), 50 mM KCl. The tubes were incubated at 16°C for 30 minutes with regular mixing. The mixture was then transferred to a spin column to separate solid from liquid phases.

Samples were finally eluted from the Glutathione Sepharose

4B matrix for 10 minutes at room temperature in 100 µl of 10 mM reduced glutathione in 50 mM tris-HCl (pH 8.0).

Capture of biotinylated PCR product strands and release of non-biotinylated strands

All separations were carried out using an Amersham magnetic separator (RPN1682, batch #1).

100 µl of the eluate from the N62D T4 endonuclease VII mismatch-capture reaction, and the pre-enrichment control were each mixed with an equal volume of 4 mg/ml streptavidin-coated colloidal Fe_3O_4 particles in 20 mM tris-HCl (pH 7.4), 2 mM EDTA (pH 8.0), 2 M NaCl. The tubes were incubated at room temperature for 30 minutes with regular mixing.

The streptavidin-coated colloidal Fe_3O_4 particles were then washed twice with 500 µl of 10 mM tris-HCl (pH 7.4), 1 mM EDTA (pH 8.0), 1 M NaCl at room temperature.

The washed streptavidin-coated colloidal Fe_3O_4 particles were incubated in 10 µl of 0.1 M NaOH for 10 minutes at room temperature. The supernatant was removed and added to 2.5 µl of 2 M HEPES (free acid).

5 µl fractions were finally spotted onto Hybond N+ nylon membranes along with 200 ng, 20 ng, 2 ng, and 200 pg amounts of the following:

C = non-biotinylated #1M single-stranded DNA (the lower of the two #1M strands shown previously)
D = non-biotinylated #7C single-stranded DNA (the lower of the two #7C strands shown previously)

Probe labelling

Probe labelling was exactly as described in example #3a.

Hybridization

Each membrane was placed in a 55 mm x 35 mm x 21 mm plastic box and 2.5 ml of pre-hybridization solution (5x SSC; Denhardt's solution; 1 % SDS; 10 % dextran sulphate [MW 500,000]; 0.3 % tetrasodium pyrophosphate; 100 µg/ml denatured, sonicated DNA – pre-warmed to 42°C) was added. Each box was closed and incubated at 42°C for 10 minutes on a rocking platform. The pre-hybridization solution was removed and replaced with hybridization solution (5x SSC; Denhardt's solution; 1 % SDS; 10 % dextran sulphate [MW 500,000]; 0.3 % tetrasodium pyrophosphate; 100 µg/ml denatured, sonicated DNA – containing 2.5 µl of the appropriate ^{33}P-labelled probe) and the box was incubated at 42°C overnight on a rocking platform.

Washing

The membranes were drained and transferred to 200 ml of 2x SSC, 0.1 % SDS at 42°C for 10 minutes. A further wash was carried out in 0.2x SSC, 0.1 % SDS at 42°C for 10 minutes. The membranes were rinsed in 2x SSC at room temperature and laid out on blotting paper to remove excess liquid. Once dry, the membrane was covered in Saran Wrap and exposed to a Kodak Phosphor Screen for 1 hour. The phosphor screen was subsequently imaged using a Molecular Dynamics Storm 860 Phosphorimager.

Signal intensities from the spots above were again reported using ImageQuant 5.0 software (Molecular Dynamics), with the SumAboveBG figures being used after drawing a grid over the array of spots.

#1 probe SumAbov BG signals

200 ng #1M ssDNA 22418970	20 ng #1M ssDNA 3317389	2 ng #1M ssDNA 270664	200 pg #1M ssDNA 25763	GSH eluted sample 144398
200 ng #7C ssDNA 229951	20 ng #7C ssDNA 105343	2 ng #7C ssDNA 16354	200 pg #7C ssDNA 3632	Pre-enrichment control 997077

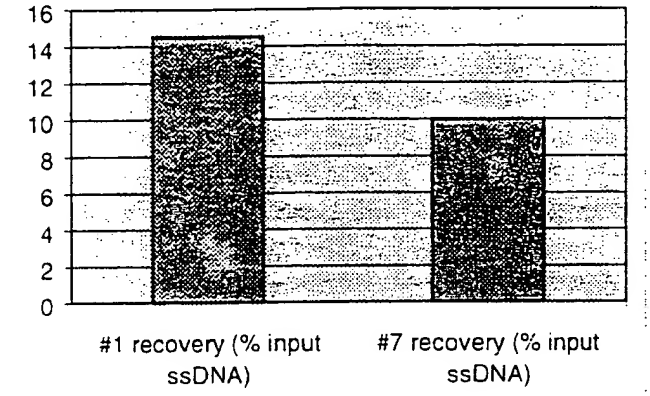
#7 probe SumAboveBG signals

200 ng #1M ssDNA 117267	20 ng #1M ssDNA 42226	2 ng #1M ssDNA 9421	200 pg #1M ssDNA 2812	GSH eluted sample 113608
200 ng #7C ssDNA 3665967	20 ng #7C ssDNA 22248096	2 ng #7C ssDNA 203869	200 pg #7C ssDNA 19406	Pre-enrichment control 1135786

Recovery

Elution conditions	#1 signal (% input ssDNA)	#7 signal (% input ssDNA)	#1/#7 signal ratio
10 mM reduced glutathione in 50 mM tris-HCl (pH 8.0)	14.5 %	10.0 %	1.45

The enrichment results are presented graphically below:



Dedicated to the memory of Chris Griffin and Richard Beer